

# Ergodicity of observation-driven time series models and consistency of the maximum likelihood estimator

R. Douc<sup>a</sup>, P. Doukhan<sup>b</sup>, E. Moulines<sup>c</sup>

<sup>a</sup>*Department CITI, CNRS UMR 5157, Telecom Sudparis, Evry. France.*

<sup>b</sup>*Department of mathematics, University of Cergy-Pontoise, France.*

<sup>c</sup>*Department LTCI, CNRS UMR 5141, Telecom Paristech, Paris. France.*

---

## Abstract

This paper deals with a general class of observation-driven time series models with a special focus on time series of counts. We provide conditions under which there exist strict-sense stationary and ergodic versions of such processes. The consistency of the maximum likelihood estimators is then derived for well-specified and misspecified models.

*Keywords:* consistency, ergodicity, time series of counts, maximum likelihood, observation-driven models, stationarity.

---

There has recently been a strong renewed interest in developing models for time series of counts which arise in a wide variety of applications: economics, finance, epidemiology, population dynamics... Among the models proposed so far, observation-driven models introduced by Cox (1981) plays an important role (see (Kedem and Fokianos, 2002, Chapter 4) for a comprehensive account and Tjøstheim (2012) for a recent survey). In time series of counts, the observations are the realisations of some integer-valued distribution (e.g. Poisson, negative binomial, ...) depending on some parameters that drives the dynamic of the model. In this paper, we focus on the so-called observation-driven time series models in which the parameter depends solely the past observations. Examples of such models include Poisson integer-valued GARCH (INGARCH) (see Ferland et al. (2006) or Zhu (2012), Fokianos et al. (2009)), Poisson threshold models (see Henderson et al. (2011)), log-linear Poisson autoregression (see Fokianos and Tjøstheim (2011)); see also Davis et al. (2003), Davis and Liu (2012) and Neumann (2011) for other observation-driven models for Poisson counts.

This paper discusses the theory and inference for a general class of observation-driven models which includes the models introduced above as particular examples. Compared to the approach introduced in Fokianos and Tjøstheim (2011), our argument is not based on the so-called perturbation technique. Recall that

---

*Email addresses:* `randal.douc@telecom-sudparis.eu` (R. Douc), `doukhan@u-cergy.fr` (P. Doukhan), `moulines@telecom-paristech.fr` (E. Moulines)

this technique consists in two steps: in a first step, a perturbed version of the process is shown to be geometrically ergodic, in a second step, the perturbed process is shown to converge to the original one by letting the perturbation goes to 0. These two steps make it possible to develop a likelihood theory on the perturbed process and then to take the limit. As argued by Doukhan (2012), this approximation technique might seem unnatural and is technically involved. In addition, it heavily relies on the Poisson assumption. The approach developed by Neumann (2011) is more direct but is based on a contraction assumption on the intensity of the Poisson variable which is not satisfied, for example neither in the log-linear Poisson autoregression model nor in the Poisson threshold model. We do not follow the weak dependence approach which as outlined in Doukhan et al. (2012) also implies unnecessary Lipschitz assumptions of the model and does not yield directly a theory for likelihood inference. Those authors apply Doukhan and Wintenberger (2008) results; the latter use a contraction argument also adapted to deal with more general infinite memory models which essentially extends on assumptions (13) below relative to the current Markov case. Those authors also derived weak dependence conditions for such models; we should anyway quote that such Taylor-made dependence conditions do not allow as performing results as the present techniques.

Our approach is based on the theory of Markov chains without irreducibility assumption. We first prove the existence of a stationary distribution using the result of Tweedie (1988). The main difficulty when the Markov chain is not necessarily irreducible consists in proving the uniqueness of the stationary distribution. For that purpose, we extend the delicate argument introduced by Henderson et al. (2011) and based on the theory of asymptotically strong Feller Markov chains (see Hairer and Mattingly (2006)). Our extension introduces a drift term which adds considerable flexibility on the model assumptions and allow to cover the log-linear Poisson autoregression model under assumptions which are weaker than those reported in Fokianos and Tjøstheim (2011). We then establish ergodicity for the two-sided stationary version of the process under the sole assumption of existence and uniqueness of the stationary distribution. Finally, we develop the theory of likelihood inference by approximating the conditional likelihood by an appropriately defined stationary version of it, which is shown to converge using classical ergodic theory arguments. Our likelihood inference theory covers both well-specified and misspecified models. We focus on the consistency of the conditional likelihood estimator but the asymptotic normality can also be covered using stationary martingale arguments. Due to space constraints, this will be reported in a forthcoming paper.

The organization of the paper is as follows. Section 1 formulates the model, establishes the existence and uniqueness of the invariant distribution and shows the ergodicity and existence of some moments for the observation process. The maximum likelihood estimates of the parameters and the relevant asymptotic theory are then derived in Section 2. Examples of threshold autoregressive and log Poisson counts are used to illustrate our findings. The proofs are given in Section 3. Finally, the Appendix contains general statements about the ergodicity of Markov chains under minimal assumptions which might be of independent

interest.

## 1. Ergodicity of the Observation-driven time series model

Let  $(\mathsf{X}, d)$  be a locally compact, complete and separable metric space and denote by  $\mathcal{X}$  the associated Borel sigma-field. Let  $(\mathsf{Y}, \mathcal{Y})$  be a measurable space,  $H$  a Markov kernel from  $(\mathsf{X}, \mathcal{X})$  to  $(\mathsf{Y}, \mathcal{Y})$  and  $(x, y) \mapsto f_y(x)$  a measurable function from  $(\mathsf{X} \times \mathsf{Y}, \mathcal{X} \otimes \mathcal{Y})$  to  $(\mathsf{X}, \mathcal{X})$ .

**Definition 1.** An observation-driven time series model on  $\mathbb{N}$  is a stochastic process  $\{(X_n, Y_n), n \in \mathbb{N}\}$  on  $\mathsf{X} \times \mathsf{Y}$  satisfying the following recursions: for all  $k \in \mathbb{N}$ ,

$$\begin{aligned} Y_{k+1} | \mathcal{F}_k &\sim H(X_k; \cdot) , \\ X_{k+1} &= f_{Y_{k+1}}(X_k) , \end{aligned} \tag{1}$$

where  $\mathcal{F}_k = \sigma(X_\ell, Y_\ell; \ell \leq k, \ell \in \mathbb{N})$ . Similarly,  $\{(X_n, Y_n), n \in \mathbb{Z}\}$  is an observation-driven time series model on  $\mathbb{Z}$  if the previous recursion holds for all  $k \in \mathbb{Z}$  with  $\mathcal{F}_k = \sigma(X_\ell, Y_\ell; \ell \leq k, \ell \in \mathbb{Z})$ .

Observation-driven time series models have been introduced by Cox (1981) and later considered by Streett (2000), Davis et al. (2003), Fokianos et al. (2009), Neumann (2011) and Doukhan et al. (2012).

In an *observation-driven time series* model,  $\{Y_n\}_{n \in \mathbb{N}}$  are observed whereas  $\{X_n\}_{n \in \mathbb{N}}$  are not observed. This model shares similarities with Hidden Markov Models, the main difference lying in the fact that given  $X_0$  and  $k$  successive observations  $Y_0, \dots, Y_k$ , (1) allows to compute  $X_k$ . In the following, the notation  $u_{s:t}$  stands for  $(u_s, \dots, u_t)$  for  $s \leq t$ .

**Example 2.** The *GARCH(1,1)* model defined by

$$\begin{aligned} Y_{k+1} | \sigma_{0:k}^2, Y_{0:k} &\sim \mathcal{N}(0, \sigma_k^2) , \\ \sigma_{k+1}^2 &= d + a\sigma_k^2 + bY_{k+1}^2 , \end{aligned}$$

where  $\min(d, a, b) > 0$  can be written as in (1) by setting  $X_k = \sigma_k^2$  and  $f_y(x) = d + ax + by^2$ .

**Example 3.** The *Poisson threshold* model defined by

$$\begin{aligned} Y_{k+1} | X_{0:k}, Y_{0:k} &\sim \mathcal{P}(X_k) , \\ X_{k+1} &= \omega + aX_k + bY_{k+1} + (cX_k + dY_{k+1}) \mathbb{1}\{Y_{k+1} \notin (L, U)\} , \end{aligned}$$

where  $\mathcal{P}(\lambda)$  is the Poisson distribution of parameter  $\lambda$  and  $0 < L < U < \infty$  can be written as in (1) by setting  $f_y(x) = \omega + ax + by + (cx + dy) \mathbb{1}\{y \notin (L, U)\}$ .

Note that  $X_n$  being the parameter of a Poisson distribution, it should be nonnegative. It is therefore usually assumed that  $X_0 \geq \omega$  and  $\min(\omega, a, b, a + c, b + d) > 0$ .

**Example 4.** The log-linear Poisson autoregression model introduced by Fokianos and Tjøstheim (2011) and defined by

$$\begin{aligned} Y_{k+1} | X_{0:k}, Y_{0:k} &\sim \mathcal{P}(e^{X_k}) , \\ X_{k+1} &= d + aX_k + b \ln(1 + Y_{k+1}) , \end{aligned}$$

where  $\mathcal{P}(\lambda)$  is the Poisson distribution of parameter  $\lambda$  can also be written as in (1) by setting  $f_y(x) = d + ax + b \ln(1 + y)$ .

A natural question is to find conditions under which there exists a strict-sense stationary and ergodic version of the observation process  $\{Y_k\}_{k \in \mathbb{N}}$ . Note For the GARCH(1,1) model as described in Example 2, this problem can be easily solved by exploiting known results on random coefficient autoregressive processes; see for example Brandt (1986), Bougerol and Picard (1992) and the references therein.

Since  $\{Y_k\}_{k \in \mathbb{N}}$  is not itself a Markov chain, a classical approach is to prove the existence of a strict-sense stationary ergodic process  $\{Y_k\}_{k \in \mathbb{N}}$  as a deterministic function of an ergodic Markov chain. To this aim, it is worthwhile to note that  $\{((X_n, Y_n), \mathcal{F}^{X,Y}), n \in \mathbb{N}\}$  is a Markov chain on  $(\mathbf{X} \times \mathbf{Y}, \mathcal{X} \otimes \mathcal{Y})$  with respect to its natural filtration

$$\mathcal{F}^{X,Y} = (\mathcal{F}_k^{X,Y}, k \in \mathbb{N}), \quad \text{where} \quad \mathcal{F}_\ell^{X,Y} = \sigma((X_k, Y_k), 1 \leq k \leq \ell, X_0),$$

and that  $\{(X_n, \mathcal{F}^X), n \in \mathbb{N}\}$  is also a Markov chain on  $(\mathbf{X}, \mathcal{X})$  with respect to its natural filtration

$$\mathcal{F}^X = (\mathcal{F}_k^X, k \in \mathbb{N}), \quad \text{where} \quad \mathcal{F}_k^X = \sigma(X_\ell, 0 \leq \ell \leq k).$$

Denote now by  $Q$  the Markov kernel associated to  $\{X_k, k \in \mathbb{N}\}$  defined implicitly by the recursions (1).

In this section, we derive general conditions expressed in terms of  $H$  and  $f$  under which  $\{X_k, k \in \mathbb{N}\}$  and  $\{(X_k, Y_k), k \in \mathbb{N}\}$  admits a unique invariant probability distribution. This is a particularly tricky task when the observation process  $\{Y_n\}_{n \in \mathbb{N}}$  is integer-valued as in Example 3 and Example 4. In such case, the Markov chain  $\{X_n\}_{n \in \mathbb{N}}$  takes value on

$$\{f_{y_k} \circ \dots \circ f_{y_1}(x_0) : k \in \mathbb{N}, (y_1, \dots, y_k) \in \mathbb{Z}^k\},$$

which is a countable subset of  $\mathbf{X}$ . When starting from two different points  $x_0$  and  $x'_0$ , the values taken by  $\{X_n\}_{n \in \mathbb{N}}$  may belong to two disjoint countable subsets of  $\mathbf{X}$ . In that case, the total variation distance between  $Q^n(x_0, \cdot)$  and  $Q^n(x'_0, \cdot)$  is always equal to 2 regardless the values of  $n \in \mathbb{N}$  and thus does not converge to 0. We therefore stress that the results obtained in the sequel do not assume that the Markov chain is irreducible.

### 1.1. Coupling construction and main results

The proof is based on a coupling construction on Markov chains which is now described.

Introduce a kernel  $\bar{H}$  from  $(\mathbf{X}^2, \mathcal{X}^{\otimes 2})$  to  $(\mathbf{Y}^2, \mathcal{Y}^{\otimes 2})$  satisfying the following conditions on the marginals: for all  $(x, x') \in \mathbf{X}^2$  and  $A \in \mathcal{Y}$ ,

$$\bar{H}((x, x'); A \times \mathbf{Y}) = H(x, A), \quad \bar{H}((x, x'); \mathbf{Y} \times A) = H(x', A). \quad (2)$$

Let  $\mathbf{C} \in \mathcal{Y}^{\otimes 2}$  such that  $\bar{H}((x, x'); \mathbf{C}) \neq 0$  and consider the Markov chain  $\{Z_k = (X_k, X'_k, U_k), n \in \mathbb{N}\}$  on the "extended" space  $(\mathbf{X}^2 \times \{0, 1\}, \mathcal{X}^{\otimes 2} \otimes \mathcal{P}(\{0, 1\}))$  with transition kernel  $\bar{Q}$  implicitly defined as follows. Given  $Z_k = (x, x', u) \in \mathbf{X}^2 \times \{0, 1\}$ , draw  $(Y_{k+1}, Y'_{k+1})$  according to  $\bar{H}((x, x'); \cdot)$  and set

$$X_{k+1} = f_{Y_{k+1}}(x), \quad X'_{k+1} = f_{Y'_{k+1}}(x'), \quad U_{k+1} = \mathbb{1}_{\mathbf{C}}(Y_{k+1}, Y'_{k+1}), \\ Z_{k+1} = (X_{k+1}, X'_{k+1}, U_{k+1}).$$

The conditions on the marginals of  $\bar{H}$ , given by (2) also imply conditions on the marginals of  $\bar{Q}$ : for all  $A \in \mathcal{X}$  and  $z = (x, x', u) \in \mathbf{X}^2 \times \{0, 1\}$ ,

$$\bar{Q}(z; A \times \mathbf{X} \times \{0, 1\}) = Q(x; A), \quad \bar{Q}(z; \mathbf{X} \times A \times \{0, 1\}) = Q(x'; A). \quad (3)$$

For  $z = (x, x', u) \in \mathbf{X}^2 \times \{0, 1\}$ , write

$$\alpha(x, x') = \bar{Q}(z; \mathbf{X}^2 \times \{1\}) = \bar{H}((x, x'); \mathbf{C}) \neq 0. \quad (4)$$

The quantity  $\alpha(x, x')$  is thus the probability of the event  $\{U_1 = 1\}$  conditionally on  $Z_0$ , taken on  $Z_0 = z$ . Denote by  $Q^\#$  the kernel on  $(\mathbf{X}^2, \mathcal{X}^{\otimes 2})$  defined by: for all  $z = (x, x', u) \in \mathbf{X}^2 \times \{0, 1\}$  and  $A \in \mathcal{X}^{\otimes 2}$ ,

$$Q^\#((x, x'); A) = \frac{\bar{Q}(z; A \times \{1\})}{\bar{Q}(z; \mathbf{X}^2 \times \{1\})},$$

so that using (4),

$$\bar{Q}(z; A \times \{1\}) = \alpha(x, x') Q^\#((x, x'); A). \quad (5)$$

This shows that  $Q^\#((x, x'); \cdot)$  is the distribution of  $(X_1, X'_1)$  conditionally on  $(X_0, X'_0, U_1) = (x, x', 1)$ . Consider the following assumptions:

**(A1)** The Markov kernel  $Q$  is weak Feller. Moreover, there exist a compact set  $C \in \mathcal{X}$ ,  $(b, \epsilon) \in \mathbb{R}_*^+ \times \mathbb{R}_*^+$  and a function  $V : \mathbf{X} \rightarrow \mathbb{R}^+$  such that

$$QV \leq V - \epsilon + b\mathbb{1}_C. \quad (6)$$

Following (Meyn and Tweedie, 1993, Definition 6.1.2), a point  $x_0 \in \mathbf{X}$  is said to be *reachable* for the Markov kernel  $Q$  if for all  $x \in \mathbf{X}$  and all open sets  $A$  containing  $x_0$ , we have  $\sum_n Q^n(x, A) > 0$ .

**(A2)** The Markov kernel  $Q$  has a reachable point.

In what follows, if  $(\mathbf{E}, \mathcal{E})$  a measurable space,  $\xi$  a probability distribution on  $(\mathbf{E}, \mathcal{E})$  and  $R$  a Markov kernel on  $(\mathbf{E}, \mathcal{E})$ , we denote by  $\mathbb{P}_\xi^R$  the probability induced on  $(\mathbf{E}^\mathbb{N}, \mathcal{E}^{\otimes \mathbb{N}})$  by a Markov chain with transition kernel  $R$  and initial distribution  $\xi$ . We denote by  $\mathbb{E}_\xi^R$  the associated expectation.

(A3) There exist a kernel  $\bar{Q}$  on  $(X^2 \times \{0, 1\}, \mathcal{X}^2 \otimes \mathcal{P}(\{0, 1\}))$ , a kernel  $Q^\sharp$  on  $(X^2, \mathcal{X}^{\otimes 2})$  and a measurable function  $\alpha : X^2 \rightarrow \{0, 1\}$  satisfying (3) and (5), a measurable function  $W : X^2 \rightarrow [1, \infty)$  and real numbers  $(D, \zeta_1, \zeta_2, \rho) \in (\mathbb{R}^+)^3 \times (0, 1)$  such that for all  $(x, x') \in X^2$ ,

$$1 - \alpha(x, x') \leq d(x, x')W(x, x') \quad (7)$$

$$\mathbb{E}_{\delta_x \otimes \delta_{x'}}^{Q^\sharp} [d(X_n, X'_n)] \leq D\rho^n d(x, x') \quad (8)$$

$$\mathbb{E}_{\delta_x \otimes \delta_{x'}}^{Q^\sharp} [d(X_n, X'_n)W(X_n, X'_n)] \leq D\rho^n d^{\zeta_1}(x, x')W^{\zeta_2}(x, x') \quad (9)$$

Moreover, for all  $x \in X$ , there exists  $\gamma_x > 0$  such that

$$\sup_{x' \in B(x, \gamma_x)} W(x, x') < \infty. \quad (10)$$

**Remark 5.** The assumption (A1) implies by (Tweedie, 1988, Theorem 2) that the Markov kernel  $Q$  admits at least one stationary distribution. Assumptions (A2-3) are then used to show that this stationary distribution is unique.

**Remark 6.** These assumptions weaken the Lipschitz conditions obtained by (Henderson et al., 2011, eq (15)) by introducing a "drift" function  $W$  in (7). This allows to treat for example the Log-linear Poisson autoregression under minimal assumptions. It thus answers to an open question raised by (Henderson et al., 2011, p. 816) on dealing with models which do not satisfy Lipschitz condition as expressed in (Henderson et al., 2011, eq (15)).

**Remark 7.** Eq (5) shows that we can simulate  $(X_1, X'_1, U_1)$  according to  $\bar{Q}((x, x', u); \cdot)$  as follows. Toss a coin with probability of heads  $\alpha(x, x')$ . If the coin lands head, then set  $U_1 = 1$  and draw  $(X_1, X'_1) \sim Q^\sharp((x, x'); \cdot)$ . Otherwise, set  $U_1 = 0$  and draw  $(X_1, X'_1)$  according to

$$A \mapsto \frac{\bar{Q}((x, x', u); A \times \{0\})}{1 - \alpha(x, x')}. \quad (11)$$

Under (8) and (9), the stochastic processes

$$\{d(X_k, X'_k), k \in \mathbb{N}\}, \quad \text{and} \quad \{d(X_k, X'_k)W(X_k, X'_k), k \in \mathbb{N}\},$$

conditionally on the fact that the coin lands heads repeatedly, goes geometrically fast to 0 in expectation. When the coin lands tail, nothing is assumed about the behavior of these processes but we can bound the probability of this event by (7).

**Theorem 8.** Assume that (A1-3) hold. Then, the Markov kernel  $Q$  admits a unique invariant probability measure.

*Proof.* The proof is postponed to Section 3.  $\square$

Note that Theorem 8 does not provide a rate of convergence to the stationary distribution. Nevertheless, when discussing inference in these models, some moment conditions with respect to the stationary distribution are needed. The following Lemma allows to assess if a function  $f$  is integrable with respect to an invariant distribution of the Markov kernel  $Q$ .

**Lemma 9.** Assume that the Markov kernel  $Q$  admits an invariant kernel  $\pi$  and that there exist a measurable function  $V : \mathsf{X} \rightarrow \mathbb{R}^+$  and real numbers  $(\lambda, \beta) \in (0, 1) \times \mathbb{R}^+$  such that  $QV \leq \lambda V + \beta$ . Then,

$$\pi V \leq \beta / (1 - \lambda) < \infty.$$

*Proof.* The proof is postponed to Section 3.  $\square$

**Proposition 10.** Assume that the Markov kernel  $Q$  admits a unique invariant probability measure. Then, there exists a strict-sense stationary ergodic process on  $\mathbb{Z}$ ,  $\{Y_n\}_{n \in \mathbb{Z}}$ , solution to the recursion (1).

*Proof.* Denote by  $\pi$  the unique invariant distribution of the Markov kernel  $Q$ . Now, let  $\{(X_n, Y_n), n \in \mathbb{N}\}$  be the Markov chain satisfying (1). If  $\bar{\pi}$  is an invariant distribution for  $\{(X_n, Y_n), n \in \mathbb{N}\}$ , then the marginal distribution  $A \mapsto \bar{\pi}(A \times \mathsf{Y})$  is a stationary distribution for the Markov kernel  $Q$  and since  $\pi$  is unique,  $\bar{\pi}(A \times \mathsf{Y}) = \pi(A)$ . If  $(X_0, Y_0) \sim \bar{\pi}$ , then by (1),  $(X_1, Y_1)$  is distributed according to  $B \mapsto \iint \pi(dx) H(x; dy_1) \mathbb{1}_B(f_{y_1}(x), y_1)$ . Since  $\bar{\pi}$  is an invariant distribution for  $\{(X_n, Y_n), n \in \mathbb{N}\}$ , we therefore obtain,

$$\bar{\pi}(B) = \iint \pi(dx) H(x; dy_1) \mathbb{1}_B(f_{y_1}(x), y_1), \quad \text{for all } B \in \mathcal{X} \otimes \mathcal{Y}. \quad (12)$$

Thus, the Markov chain  $\{(X_n, Y_n), n \in \mathbb{N}\}$  has a unique invariant distribution given by (12). By applying Theorem 32 and Theorem 33, there exists a strict-sense stationary ergodic process on  $\mathbb{Z}$ ,  $\{(X_n, Y_n), n \in \mathbb{Z}\}$ , solution to the recursion (1). The proof follows.  $\square$

We end the section by providing some practical conditions for checking (8) and (9) in (A3).

**Lemma 11.** Assume that either (i) or (ii) or (iii) (defined below) holds.

(i) There exists  $(\rho, \beta) \in (0, 1) \times \mathbb{R}$  such that for all  $(x, x') \in \mathsf{X}^2$

$$d(X_1, X'_1) \leq \rho d(x, x'), \quad \mathbb{P}_{\delta_x \otimes \delta_{x'}}^{Q^\sharp} \text{-a.s.} \quad (13)$$

$$Q^\sharp W \leq W + \beta \quad (14)$$

(ii) (8) holds and  $W$  is bounded.

(iii) (8) holds and there exists  $0 < \alpha < \alpha'$  and  $\beta \in \mathbb{R}^+$  such that for all  $(x, x') \in \mathsf{X}^2$

$$\begin{aligned} d(x, x') &\leq W^\alpha(x, x') \\ Q^\sharp W^{1+\alpha'} &\leq W^{1+\alpha'} + \beta \end{aligned}$$

Then, (8) and (9) hold.

**Remark 12.** *Ergodicity under Lipschitz conditions have been studied in a wide literature including Sunyach (1975) or Diaconis and Freedman (1999), but the fact that the various contraction conditions in Lemma 11 are related to the kernel  $Q^\sharp$  and not to the kernels  $Q$  or  $\bar{Q}$  make it possible to check the assumptions (8) and (9) quite directly.*

*Proof.* See Section 3. □

## 1.2. Examples

### 1.2.1. A Poisson threshold model

Existence and uniqueness of the stationary distribution for the Poisson threshold model have been already discussed in Henderson et al. (2011). We can obtain the same results by applying Theorem 8 provided that assumptions (A1-3) hold. Consider a Markov chain  $\{X_n, n \in \mathbb{N}\}$  with a transition kernel  $Q$  given implicitly by the following recursive equations:

$$\begin{aligned} Y_{n+1} | X_{0:n}, Y_{0:n} &\sim \mathcal{P}(X_n), \\ X_{n+1} &= \omega + aX_n + bY_{n+1} + (cX_n + dY_{n+1}) \mathbb{1}\{Y_{n+1} \notin (L, U)\}, \end{aligned}$$

where  $0 < L < U < \infty$ . Moreover, to keep the parameter of the Poisson distribution positive, it is assumed that  $X_0 \geq \omega$  and  $\min(\omega, a, b, a + c, b + d) > 0$ . Here, we set  $\mathbf{X} = \mathbb{R}^+$ ,  $d(x, x') = |x - x'|$  and

$$f_y(x) = \omega + ax + by + (cx + dy) \mathbb{1}\{y \notin (L, U)\}.$$

**Lemma 13.** *Assume that  $a \vee (a + c) < 1$ , then (A3) holds.*

*Proof.* Define implicitly  $\bar{Q}$  as the transition kernel Markov chain  $\{Z_n, n \in \mathbb{N}\}$  with  $Z_n = (X_n, X'_n, U_n)$  in the following way. Given  $Z_n = (x, x', u)$ , if  $x \leq x'$ , draw independently  $Y_{n+1} \sim \mathcal{P}(x)$ ,  $V_{n+1} \sim \mathcal{P}(x' - x)$  and set  $Y'_{n+1} = Y_{n+1} + V_{n+1}$ . Otherwise, draw independently  $Y'_{n+1} \sim \mathcal{P}(x')$  and  $V_{n+1} \sim \mathcal{P}(x - x')$  and set  $Y_{n+1} = Y'_{n+1} + V_{n+1}$ . In all cases, set

$$\begin{aligned} X_{n+1} &= f_{Y_{n+1}}^\theta(x), \\ X'_{n+1} &= f_{Y'_{n+1}}^\theta(x'), \\ U_{n+1} &= \mathbb{1}\{Y_{n+1} = Y'_{n+1}\} = \mathbb{1}\{V_{n+1} = 0\}, \\ Z_{n+1} &= (X_{n+1}, X'_{n+1}, U_{n+1}). \end{aligned}$$

Note again that if  $Y \sim \mathcal{P}(\lambda)$ ,  $V \sim \mathcal{P}(\lambda')$  and  $(Y, V)$  are independent, then  $Y + V \sim \mathcal{P}(\lambda + \lambda')$ . This implies that  $\bar{Q}$  satisfies the marginal conditions (3). Define for all  $x^\sharp = (x, x') \in \mathbb{R}^2$ ,  $Q^\sharp(x^\sharp, \cdot)$  as the law of  $(X_1, X'_1)$  where

$$X_1 = f_Y^\theta(x), \quad X'_1 = f_{Y'}^\theta(x'),$$

and  $Y \sim \mathcal{P}(x \wedge x')$ , and set, for all  $x^\sharp = (x, x') \in \mathbb{R}^2$ ,

$$\alpha(x^\sharp) = \exp\{-|x - x'|\}.$$



With these definitions, obviously,  $\bar{Q}$ ,  $\alpha$  and  $Q^\sharp$  satisfy (5). Moreover, using  $1 - e^{-u} \leq u$ , we obtain

$$1 - \alpha(x^\sharp) = 1 - \exp\{-|x - x'|\} \leq |x - x'|.$$

so that (7) holds with  $W = \mathbb{1}_{\mathbb{R}^2}$ . To obtain (8) and (9), we apply Lemma 11 by checking (i) in Lemma 11.

$$\mathbb{P}_{\delta_x \otimes \delta_{x'}}^{Q^\sharp} \{|X_1 - X'_1| = |a + c\mathbb{1}\{Y_1 \notin (L, U)\}||x - x'| \leq \rho|x - x'|\} = 1, \quad (15)$$

where  $\rho = a \vee (a + c) < 1$ . The function  $W$  being constant, (i) holds and the proof is completed.  $\square$

**Proposition 14.** *Assume that  $(a + b + c + d) \vee a < 1$ , then the Markov kernel  $Q$  admits a unique stationary distribution  $\pi$ . Moreover,  $\pi V < \infty$  where  $V$  is the function  $V : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  defined by  $V(x) = x$ .*

*Proof.* According to Theorem 8 and Lemma 13, it is enough to show (A1-2) to obtain the existence and unicity of an invariant probability measure  $\pi$ . We start with (A1). A random variable of distribution  $\mathcal{P}(\lambda)$  converges weakly to a random variable of distribution  $\mathcal{P}(\lambda')$  as  $\lambda \rightarrow \lambda'$ . This implies by Slutsky's Lemma that if  $N : \mathbb{R}_+ \rightarrow \mathbb{N}$ ,  $x \mapsto N(x)$  is a Poisson process of unit intensity, then

$$X_1(x) = \omega + ax + bN(x) + (cx + dN(x))\mathbb{1}\{N(x) \notin (L, U)\}$$

converges weakly to  $X_1(x')$  as  $x \rightarrow x'$ . Therefore,  $Q$  is weakly Feller. Moreover, it can be readily checked that the nonnegative function  $V(x) = x$  ( $V$  is indeed nonnegative as a function defined on  $\mathbf{X} = \mathbb{R}^+$ ) satisfies:

$$QV(x) = (a + b + c\mathbb{P}[N(x) \notin (L, U)] + d\mathbb{E}[N(x)\mathbb{1}_{N(x) \notin (L, U)}]/x)V(x) + \omega.$$

It can be easily checked that

$$\lim_{x \rightarrow \infty} \mathbb{P}[N(x) \notin (L, U)] = 1, \quad \text{and} \quad \lim_{x \rightarrow \infty} \mathbb{E}[N(x)\mathbb{1}_{N(x) \notin (L, U)}]/x = 1,$$

so that

$$\lim_{x \rightarrow \infty} \frac{QV(x)}{V(x)} = a + b + c + d < 1, \quad \text{and} \quad \sup_{0 \leq x \leq M} QV(x) < \infty, \quad \forall M \in \mathbb{R}^+.$$

These two properties imply that there exist  $(\lambda, \beta) \in (0, 1) \times \mathbb{R}^+$  such that

$$QV \leq \lambda V + \beta. \quad (16)$$

Thus, the drift condition (6) holds. Thus, (A1) is satisfied. Set  $x_\infty = \omega/(1 - a - c)$  and let  $C$  be an open set containing  $x_\infty$ . Let  $x \in \mathbb{R}$  and define recursively the sequence  $x_0 = x$  and for all  $k \geq 1$ ,  $x_k = \omega + (a + c)x_{k-1}$ . Since  $(a + c) < 1$ , this

sequence has a unique limiting point,  $\lim_{n \rightarrow \infty} x_n = x_\infty$ . Therefore, there exists some  $n$  such that for all  $k \geq n$ ,  $x_k \in C$ . For such  $n$ , we have

$$\begin{aligned} Q^n(x, C) &= \mathbb{P}_{\delta_x}^Q(X_n \in C) \geq \mathbb{P}_{\delta_x}^Q(X_n \in C, Y_1 = \dots = Y_n = 0) \\ &= \mathbb{P}_{\delta_x}^Q(Y_1 = \dots = Y_n = 0) > 0. \end{aligned}$$

so that (A2) holds. Moreover, since the function  $V(x) = x$  satisfies (16), Lemma 9 shows that  $\pi V < \infty$ .  $\square$

**Remark 15.** In the proof of Lemma 13, we check (A3) by verifying Lemma 11-(i). In some models, applying Lemma 11-(ii) may provide more flexibility as can be seen in the following example:

$$\begin{aligned} Y_{n+1}|X_{0:n}, Y_{0:n} &\sim \mathcal{P}(X_n), \\ X_{n+1} &= \omega + (a + c\mathbb{1}_{Y_{n+1}=0})X_n + bY_{n+1} + dY_{n+1}\mathbb{1}_{Y_{n+1}=0}, \end{aligned}$$

This is a particular Poisson threshold model as defined in Example 3 with  $(L, U) = (1/2, \infty)$ . It is assumed that  $X_0 \geq \omega$  and  $\min(\omega, a, b, a + c, b + d) > 0$ . Now, according to Lemma 13, (A3) holds if  $a \vee (a + c) < 1$ . We can now prove that (A3) holds even if  $a + c > 1$  provided that  $a \vee (a + ce^{-\omega}) < 1$ . To see this, we just adapt the proof of Lemma 13 by replacing (15) by

$$\mathbb{E}_{\delta_x \otimes \delta_{x'}}^{Q^\sharp}(|X_1 - X'_1|) = \mathbb{E}_{\delta_x \otimes \delta_{x'}}^{Q^\sharp}(a + c\mathbb{1}\{Y_1 = 0\})|x - x'| \leq \rho|x - x'|$$

where  $\rho := a + ce^{-\omega} < 1$ . This implies that (8) holds so that condition Lemma 11-(ii) holds and thus Lemma 11 concludes the proof.

### 1.2.2. Log-linear Poisson autoregression

Consider a Markov chain  $\{X_n, n \in \mathbb{N}\}$  with a transition kernel  $Q$  given implicitly by the following recursive equations:

$$\begin{aligned} Y_{n+1}|X_{0:n}, Y_{0:n} &\sim \mathcal{P}(e^{X_n}), \\ X_{n+1} &= d + aX_n + b \ln(Y_{n+1} + 1), \end{aligned} \tag{17}$$

where  $\mathcal{P}(\lambda)$  is a Poisson distribution with parameter  $\lambda$ . In this case, the state space is  $\mathbf{X} = \mathbb{R}$  which is equipped with the euclidean distance  $d(x, x') = |x - x'|$  and the function  $f_y$  is defined by:  $f_y(x) = d + ax + b \ln(1 + y)$ .

**Lemma 16.** If  $|a + b| \vee |a| \vee |b| < 1$ , then (A3) holds.

*Proof.* Define implicitly  $\bar{Q}$  as the transition kernel Markov chain  $\{Z_n, n \in \mathbb{N}\}$  with  $Z_n = (X_n, X'_n, U_n)$  in the following way. Given  $Z_n = (x, x', u)$ , if  $x \leq x'$ , draw independently  $Y_{n+1} \sim \mathcal{P}(e^x)$  and  $V_{n+1} \sim \mathcal{P}(e^{x'} - e^x)$  and set  $Y'_{n+1} = Y_{n+1} + V_{n+1}$ . Otherwise, draw independently  $Y'_{n+1} \sim \mathcal{P}(e^{x'})$  and  $V_{n+1} \sim \mathcal{P}(e^x - e^{x'})$  and set  $Y_{n+1} = Y'_{n+1} + V_{n+1}$ . In all cases, set

$$\begin{aligned} X_{n+1} &= d + ax + b \ln(Y_{n+1} + 1), \\ X'_{n+1} &= d + ax' + b \ln(Y'_{n+1} + 1), \\ U_{n+1} &= \mathbb{1}\{Y_{n+1} = Y'_{n+1}\} = \mathbb{1}\{V_{n+1} = 0\}, \\ Z_{n+1} &= (X_{n+1}, X'_{n+1}, U_{n+1}). \end{aligned}$$

Along the same lines as above,  $\bar{Q}$  satisfies the marginal conditions (3). Moreover, define for all  $x^\# = (x, x') \in \mathbf{X}^2$ ,  $Q^\#(x^\#, \cdot)$  as the law of  $(X_1, X'_1)$  where

$$\begin{aligned} X_1 &= d + ax + b \ln(Y + 1) , \quad Y \sim \mathcal{P}(e^{x \wedge x'}) , \\ X'_1 &= d + ax' + b \ln(Y + 1) , \end{aligned} \quad (18)$$

and set for all  $x^\# = (x, x') \in \mathbb{R}^2$ ,

$$\alpha(x^\#) = \exp \left\{ -e^{x \vee x'} + e^{x \wedge x'} \right\} .$$

With these definitions, obviously,  $\bar{Q}$ ,  $\alpha$  and  $Q^\#$  satisfy (5). Using twice  $1 - e^{-u} \leq u$ , we obtain

$$\begin{aligned} 1 - \alpha(x^\#) &= 1 - \exp \left\{ -e^{x \vee x'} + e^{x \wedge x'} \right\} \leq e^{x \vee x'} - e^{x \wedge x'} \\ &= e^{x \vee x'} (1 - e^{-|x - x'|}) \leq W(x, x') |x - x'| . \end{aligned}$$

with  $W(x^\#) = e^{|x \vee x'|}$  so that (7) holds. To check (8) and (9), we apply Lemma 11 by checking (i) in Lemma 11. Note first that

$$\mathbb{P}_{\delta_x \otimes \delta_{x'}}^{Q^\#} \{ |X_1 - X'_1| = |a||x - x'| \} = 1 ,$$

so that (13) is satisfied. To check (14), we will show that

$$\lim_{|x \vee x'| \rightarrow \infty} \frac{Q^\# W(x, x')}{W(x, x')} = 0 . \quad (19)$$

and for all  $M > 0$ ,

$$\sup_{|x \vee x'| \leq M} Q^\# W(x, x') < \infty . \quad (20)$$

Without loss of generality, we assume that  $x \leq x'$ . Using (18), we get

$$Q^\# W(x, x') = \mathbb{E} \left( e^{|X_1 \vee X'_1|} \right) \leq \mathbb{E}(e^{|X_1|}) + \mathbb{E}(e^{|X'_1|}) . \quad (21)$$

First consider the second term of the right-hand side of (21),

$$\mathbb{E}(e^{|X'_1|}) \leq e^{|d|} \mathbb{E}(e^{|ax' + b \ln(1+Y)|}) . \quad (22)$$

Now, note that if  $u$  and  $v$  have different signs or if  $v = 0$ , then  $|u + v| \leq |u| \vee |v|$ . Otherwise,  $|u + v| = (u + v) \mathbb{1}\{v > 0\} \vee (-u - v) \mathbb{1}\{v < 0\}$ . This implies that

$$e^{|u+v|} \leq e^{|u|} + e^{|v|} + e^{u+v} \mathbb{1}\{v > 0\} + e^{-u-v} \mathbb{1}\{v < 0\} .$$

Plugging this into (22),

$$\begin{aligned} \mathbb{E}(e^{|X'_1|}) &\leq e^{|d|} \left( e^{|a||x'|} + \mathbb{E}[(1 + Y)^{|b|}] + e^{ax'} \mathbb{E}[(1 + Y)^b] \mathbb{1}\{b > 0\} \right. \\ &\quad \left. + e^{-ax'} \mathbb{E}[(1 + Y)^{-b}] \mathbb{1}\{b < 0\} \right) . \end{aligned}$$

Note that for all  $\gamma \in [0, 1]$ ,

$$\mathbb{E}[(1 + Y)^\gamma] \leq [\mathbb{E}(1 + Y)]^\gamma = (1 + e^x)^\gamma \leq 1 + e^{\gamma x} \leq 1 + e^{\gamma x'}.$$

Moreover, since  $|b| \in [0, 1]$ , we have  $b\mathbb{1}\{b > 0\} \in [0, 1]$  and  $-b\mathbb{1}\{b < 0\} \in [0, 1]$ . Therefore,

$$\begin{aligned} \mathbb{E}(e^{|X_1|}) &\leq e^{|d|} \left( e^{|a||x'|} + 1 + e^{|b||x|} + e^{ax'}(1 + e^{bx'})\mathbb{1}\{b > 0\} \right. \\ &\quad \left. + e^{-ax'}(1 + e^{-bx'})\mathbb{1}\{b < 0\} \right) \\ &\leq e^{|d|} \left( e^{|a||x'|} + 1 + e^{|b||x|} + e^{|a||x'|} + e^{|a+b||x'|} \right) \\ &\leq e^{|d|} \left( 1 + 4e^{\gamma(|x| \vee |x'|)} \right), \end{aligned}$$

where  $\gamma = |a| \vee |b| \vee |a+b| < 1$ . The first term of the right-hand side of (21) is treated as the second term by setting  $x' = x$ . We then have

$$\mathbb{E}(e^{|X_1|}) \leq e^{|d|} \left( 1 + 4e^{\gamma(|x| \vee |x'|)} \right),$$

so that using (21),

$$Q^\# W(x, x') \leq 2e^{|d|} \left( 1 + 4e^{\gamma(|x| \vee |x'|)} \right). \quad (23)$$

Since  $\gamma \in (0, 1)$  and  $W(x, x') = e^{|x| \vee |x'|}$ , (23) implies clearly (19) and (20). The proof is completed.  $\square$

**Proposition 17.** *If  $|a+b| \vee |a| \vee |b| < 1$ , the Markov kernel  $Q$  admits a unique invariant probability measure. Moreover,  $\pi V < \infty$  where  $V(x) = e^{|x|}$ .*

**Remark 18.** (Fokianos and Tjøstheim, 2011, Lemma 2.1) have obtained that the Log-linear Poisson autoregression is close to a "perturbed" ergodic Log-linear Poisson process in the case where  $a^2 + b^2 < 1$  if  $a$  and  $b$  have different signs and  $|a+b| < 1$  otherwise. In both cases, we have  $|a+b| < 1$ . In fact, if  $a^2 + b^2 < 1$ , then  $|a| \vee |b| < 1$ . Combining it with the fact that  $a \wedge b \leq a+b \leq a \vee b$  when  $a$  and  $b$  have different signs, we obtain  $|a+b| < 1$ . Our conditions thus extends conditions of Fokianos and Tjøstheim (2011) and the results obtained here address an open question raised in (Fokianos and Tjøstheim, 2011, page 566).

*Proof.* According to Theorem 8 and Lemma 16, it is enough to show (A1-2). We consider first (A1). As above,  $X_1(x) = d + ax + b \ln(1 + N(e^x))$  converges weakly to  $X_1(x')$  as  $x \rightarrow x'$ . Therefore,  $Q$  is weakly Feller. Moreover, following the lines of Lemma 16, it can be readily checked that the function  $V(x) = e^{|x|}$  satisfies:

$$QV(x) \leq e^{|d|} \left( 1 + 4e^{\gamma(|x|)} \right),$$

where  $\gamma = |a+b| \vee |a| \vee |b| < 1$ . Thus,

$$QV(x) \leq \lambda V(x) + \beta, \quad (24)$$

for some constants  $(\lambda, \beta) \in (0, 1) \times \mathbb{R}^+$  showing **(A1)**. Consider now  $x = d/(1 - a)$ . Let  $x \in \mathbb{R}$  and let  $C$  be an open set containing  $x$ . Then, by setting  $x_0 = x$  and for all  $k \geq 1$ ,  $x_k = d + ax_{k-1}$ , we have  $\lim_{n \rightarrow \infty} x_n = x$  so that there exists some  $n$  such that for all  $k \geq n$ ,  $x_k \in C$ . For such  $n$ , we have

$$\begin{aligned} Q^n(x, C) &= \mathbb{P}_{\delta_x}^Q(X_n \in C) \geq \mathbb{P}_{\delta_x}^Q(X_n \in C, Y_1 = \dots = Y_n = 0) \\ &= \mathbb{P}_{\delta_x}^Q(Y_1 = \dots = Y_n = 0) > 0. \end{aligned}$$

so that **(A2)** holds. Since (24) holds for the function  $V(x) = e^{|x|}$ , Lemma 9 shows that  $\pi V < \infty$ . The proof follows.  $\square$

## 2. Consistency of the Maximum Likelihood Estimator

### 2.1. Misspecified models

Let  $(\Theta, d)$  be a compact metric set of  $\mathbb{R}^p$ , let  $H$  be a Markov kernel from  $(X, \mathcal{X})$  to  $(Y, \mathcal{Y})$  and let  $\{(x, y) \mapsto f_y^\theta(x), \theta \in \Theta\}$  be a family of measurable functions from  $(X \times Y, \mathcal{X} \otimes \mathcal{Y})$  to  $(X, \mathcal{X})$ . Assume that all  $x \in X$ ,  $H(x; \cdot)$  is dominated by some  $\sigma$ -finite measure  $\mu$  on  $(Y, \mathcal{Y})$  and denote by  $h(x; \cdot)$  its Radon-Nikodym derivative:  $h(x; y) = dH(x; \cdot)/d\mu(y)$ . Assume that  $h(x; y) > 0$  for all  $(x, y) \in X \times Y$  and that the sequence of random variables  $\{(X_k, Y_k); k \in \mathbb{N}\}$  satisfy the following recursions

$$\begin{aligned} Y_{k+1} | \mathcal{F}_k &\sim H(X_k; \cdot), \\ X_{k+1} &= f_{Y_{k+1}}^\theta(X_k), \end{aligned} \tag{25}$$

where  $\mathcal{F}_k$  is either  $\sigma(X_{0:k}, Y_{0:k})$  or  $\sigma(X_{-\infty:k}, Y_{-\infty:k})$ , depending whether the process is defined on  $\mathbb{N}$  or  $\mathbb{Z}$ . Then, the distribution of  $(Y_1, \dots, Y_n)$  conditionally on  $X_0 = x$  has a density with respect to the product measure  $\mu^{\otimes n}$  given by

$$y_{1:n} \mapsto \prod_{k=1}^n h(f^\theta \langle y_{1:k-1} \rangle(x); y_k), \tag{26}$$

where we have used the convention  $f^\theta \langle y_{1:0} \rangle(x) = x$  and the notations

$$f^\theta \langle y_{s:t} \rangle = f_{y_t}^\theta \circ f_{y_{t-1}}^\theta \circ \dots \circ f_{y_s}^\theta, \quad s \leq t. \tag{27}$$

In this section, we study the asymptotic properties of  $\theta_{n,x}$ , the conditional Maximum Likelihood Estimator (MLE) of the parameter  $\theta$  based on the observations  $(Y_1, \dots, Y_n)$  and associated to the parametric family of likelihood functions given in (26), that is, we consider

$$\theta_{n,x} \in \operatorname{argmax}_{\theta \in \Theta} \mathbb{L}_{n,x}^\theta \langle Y_{1:n} \rangle, \tag{28}$$

where

$$\mathbb{L}_{n,x}^\theta \langle y_{1:n} \rangle := n^{-1} \ln \left( \prod_{k=1}^n h(f^\theta \langle y_{1:k-1} \rangle(x); y_k) \right). \tag{29}$$

We are especially interested here in inference for *misspecified models*, that is, we *do not assume* that the distribution of the observations belongs to the set of distributions where the maximization occurs. In particular,  $(Y_n)_{n \in \mathbb{Z}}$  are not necessarily the observation process associated to the recursion (25).

Consider the following assumptions:

**(B1)**  $\{Y_n\}_{n \in \mathbb{Z}}$  is a strict-sense stationary and ergodic stochastic process

Under **(B1)**, denote by  $\mathbb{P}_\star$  the distribution of  $\{Y_n\}_{n \in \mathbb{Z}}$  on  $(Y^\mathbb{Z}, \mathcal{Y}^\mathbb{Z})$ . Write  $\mathbb{E}_\star$  the associated expectation.

**(B2)** For all  $(x, y) \in X \times Y$ , the functions  $\theta \mapsto f_y^\theta(x)$  and  $v \mapsto h(v, y)$  are continuous.

**(B3)** There exists a family of  $\mathbb{P}_\star$ -a.s. finite random variables

$$\{f^\theta \langle Y_{-\infty:k} \rangle : (\theta, k) \in \Theta \times \mathbb{Z}\}$$

such that for all  $x \in X$ ,

- (i)  $\lim_{m \rightarrow \infty} \sup_{\theta \in \Theta} d(f^\theta \langle Y_{-m:0} \rangle(x), f^\theta \langle Y_{-\infty:0} \rangle) = 0$ ,  $\mathbb{P}_\star$ -a.s.,
- (ii)  $\mathbb{P}_\star$ -a.s.,

$$\lim_{k \rightarrow \infty} \sup_{\theta \in \Theta} |\ln h(f^\theta \langle Y_{1:k-1} \rangle(x); Y_k) - \ln h(f^\theta \langle Y_{-\infty:k-1} \rangle; Y_k)| = 0,$$

- (iii)  $\mathbb{E}_\star \left[ \sup_{\theta \in \Theta} (\ln h(f^\theta \langle Y_{-\infty:k-1} \rangle; Y_k))_+ \right] < \infty$

In the following, we set for all  $(\theta, k) \in \Theta \times \mathbb{N}$ ,

$$\bar{\ell}^\theta \langle Y_{-\infty:k} \rangle := \ln h(f^\theta \langle Y_{-\infty:k-1} \rangle; Y_k). \quad (30)$$

**Remark 19.** When checking **(B3)**, we usually introduce  $f^\theta \langle Y_{-\infty:0} \rangle$  by showing that for all  $(\theta, x) \in \Theta \times X$ ,  $f^\theta \langle Y_{-m:-1} \rangle(x)$  converges,  $\mathbb{P}_\star$ -a.s., as  $m$  goes to infinity and that the limit does not depend on  $x$ . We can therefore denote by  $f^\theta \langle Y_{-\infty:0} \rangle$  this limit. With this definition, we then check **(B3)**(i)-(ii)-(iii).

**Remark 20.** When the observation process is integer-valued, the function  $y \rightarrow h(x; y)$  is a probability and thus, is less than one. It implies that for all  $\theta \in \Theta$ ,

$$(\bar{\ell}^\theta \langle Y_{-\infty:0} \rangle)_+ = (\ln h(f^\theta \langle Y_{-\infty:-1} \rangle; Y_0))_+ = 0.$$

Thus, **(B3)**-(iii) is satisfied.

Note that under **(B2)**,  $\theta_{n,x}$  is well-defined. The following theorem establishes the consistency of the sequence of estimators  $\{\theta_{n,x}, n \in \mathbb{N}\}$ .

**Theorem 21.** Assume **(B1-3)**. Then, for all  $x \in X$ ,

$$\lim_{n \rightarrow \infty} d(\theta_{n,x}, \Theta_\star) = 0, \quad \mathbb{P}_\star\text{-a.s.}$$

where  $\Theta_\star := \operatorname{argmax}_{\theta \in \Theta} \mathbb{E}(\bar{\ell}^\theta \langle Y_{-\infty:0} \rangle)$ .

*Proof.* The proof directly follows from Theorem 35 provided that

- (a)  $\mathbb{E}_\star[\sup_{\theta \in \Theta} (\bar{\ell}^\theta \langle Y_{-\infty:0} \rangle)_+] < \infty$ ,
- (b)  $\mathbb{P}_\star$ -a.s., the function  $\theta \mapsto \bar{\ell}^\theta \langle Y_{-\infty:0} \rangle$  is upper-semicontinuous,
- (c)  $\lim_{n \rightarrow \infty} \sup_{\theta \in \Theta} |\mathbb{L}_{n,x}^\theta \langle Y_{1:n} \rangle - \bar{\mathbb{L}}_n^\theta \langle Y_{-\infty:n} \rangle| = 0$ ,  $\mathbb{P}_\star$ -a.s. where

$$\bar{\mathbb{L}}_n^\theta \langle Y_{-\infty:n} \rangle = n^{-1} \sum_{k=1}^n \bar{\ell}^\theta \langle Y_{-\infty:k} \rangle .$$

But (a) follows from (iii), (b) follows by combining (i) and **(B2)** since a uniform limit of continuous functions is continuous and (c) is direct from (ii) and the definitions of  $\mathbb{L}_{n,x}^\theta \langle Y_{1:n} \rangle$  and  $\bar{\mathbb{L}}_n^\theta \langle Y_{-\infty:n} \rangle$ . The proof is completed.  $\square$

We end this section by providing a practical condition for checking the assumption **(B3)** when  $x \mapsto f_y^\theta(x)$  is Lipschitz.

**Lemma 22.** *Assume that there exists a measurable function  $\varrho : \mathsf{Y} \rightarrow \mathbb{R}^+$  such that for all  $(\theta, y, x, x') \in \Theta \times \mathsf{Y} \times \mathsf{X}^2$ ,*

$$d(f_y^\theta(x), f_y^\theta(x')) \leq \varrho(y) d(x, x') .$$

Moreover, assume that for all  $x \in \mathsf{X}$ ,

$$\mathbb{E}_\star \left[ \sup_{\theta \in \Theta} \ln^+ d(x, f_{Y_0}^\theta(x)) \right] < \infty, \mathbb{E}_\star(\ln^+ \varrho(Y)) < \infty, \text{ and } \mathbb{E}_\star(\ln \varrho(Y)) < 0 .$$

Then, assumption **(B3)**-(i) holds.

*Proof.* We have for all  $m \geq 0$ ,

$$d(f^\theta \langle Y_{-m:0} \rangle(x), f^\theta \langle Y_{-m:0} \rangle(y)) \leq d(x, y) \prod_{\ell=0}^m \varrho(Y_{-\ell}) \quad (31)$$

Taking  $y = f_{Y_{-m-1}}^\theta(x)$ , we obtain

$$d(f^\theta \langle Y_{-m:0} \rangle(x), f^\theta \langle Y_{-m-1:0} \rangle(x)) \leq d(x, f_{Y_{-m-1}}^\theta(x)) \prod_{\ell=0}^m [\varrho(Y_{-\ell})]$$

Now, since  $\mathbb{E}_{\theta_\star}(\ln \varrho(Y_0)) < 0$ ,  $\limsup_{m \rightarrow \infty} (\prod_{\ell=0}^m [\varrho(Y_{-\ell})])^{1/m} < 1$  and Lemma 36 implies that

$$\limsup_{m \rightarrow \infty} \left( \sup_{\theta \in \Theta} d(x, f_{Y_{-m-1}}^\theta(x)) \right)^{1/m} \leq 1 .$$

By the Cauchy root test, the series  $\sum \sup_{\theta \in \Theta} d(f^\theta \langle Y_{-m:0} \rangle(x), f^\theta \langle Y_{-m+1:0} \rangle(x))$  is convergent. This implies that  $\lim_{m \rightarrow \infty} f^\theta \langle Y_{-m:0} \rangle(x)$  exists,  $\mathbb{P}_\star$ -a.s. which does not depend on  $x$  by (31). This limit is denoted  $f^\theta \langle Y_{-\infty:0} \rangle$ . The convergence of the series also implies

$$\lim_{m \rightarrow \infty} \sup_{\theta \in \Theta} d(f^\theta \langle Y_{-m:0} \rangle(x), f^\theta \langle Y_{-\infty:0} \rangle) = 0 , \quad \mathbb{P}_\star\text{-a.s.}$$

so that **(B3)**-(i) holds.  $\square$

## 2.2. Well-specified models

In this section, we focus on well-specified models, that is,  $\{Y_n\}_{n \in \mathbb{N}}$  are assumed to be the observation process of a model defined by the recursions (25) with  $\theta = \theta_\star \in \Theta$ . In well-specified models, we stress the dependence in  $\theta_\star$  by using the notations

$$\mathbb{P}^{\theta_\star} := \mathbb{P}_\star, \quad \mathbb{E}^{\theta_\star} := \mathbb{E}_\star. \quad (32)$$

According to Section 1, to obtain (B1), we only need to check that, for all  $\theta \in \Theta$ , (A1-3) hold with  $f = f^\theta$ . If in addition, we assume that (B2-3) hold and that  $\Theta_\star = \{\theta_\star\}$ , then, Theorem 21 yields: for all  $x \in \mathbb{X}$ ,

$$\lim_{n \rightarrow \infty} \theta_{n,x} = \theta_\star, \quad \mathbb{P}^{\theta_\star}\text{-a.s.}$$

We now give conditions for having  $\Theta_\star = \{\theta_\star\}$ .

**Proposition 23.** *Let  $\{(X_k, Y_k), k \in \mathbb{Z}\}$  be a stationary stochastic process indexed by  $\mathbb{Z}$  which satisfies the recursions (25) for some  $\theta = \theta_\star \in \Theta$  with  $\mathcal{F}_k = \sigma(X_\ell, Y_\ell; \ell \leq k, \ell \in \mathbb{Z})$ . Assume that (B1-3) hold and that  $X_0 = f^{\theta_\star}(Y_{-\infty:0})$  then,  $H(X_0; \cdot)$  is the distribution of  $Y_1$  conditionally on  $\sigma(Y_\ell; \ell \leq 0)$ . If in addition,*

(a)  $x \mapsto H(x; \cdot)$  is one-to-one, i.e., if  $H(x; \cdot) = H(x'; \cdot)$ , then  $x = x'$ ,

(b)  $f^{\theta_\star}(Y_{-\infty:0}) = f^\theta(Y_{-\infty:0})$ ,  $\mathbb{P}^{\theta_\star}$ -a.s., implies that  $\theta = \theta_\star$ ,

then  $\Theta_\star = \{\theta_\star\}$ .

**Remark 24.** *Condition Proposition 23-(b) is similar as (Davis and Liu, 2012, Assumption (A5)). For the sake of clarity, we present here a self-contained proof for proving under these conditions that  $\Theta_\star = \{\theta_\star\}$ .*

*Proof.* For all  $A \in \mathcal{X}$ ,

$$\begin{aligned} \mathbb{E}^{\theta_\star} [\mathbb{1}_A(Y_1) | Y_{-\infty:0}] &= \mathbb{E}^{\theta_\star} [\mathbb{E}^{\theta_\star} [\mathbb{1}_A(Y_1) | X_0, Y_{-\infty:0}] | Y_{-\infty:0}] \\ &= \mathbb{E}^{\theta_\star} [\mathbb{E}^{\theta_\star} [\mathbb{1}_A(Y_1) | X_0] | Y_{-\infty:0}] = \mathbb{E}^{\theta_\star} [\mathbb{1}_A(Y_1) | X_0] = H(X_0; A), \end{aligned} \quad (33)$$

where we have used that  $X_0$  is  $\sigma(Y_\ell, \ell \leq 0)$ -measurable. This concludes the first part of Proposition 23. Now, for all  $\theta \in \Theta$ ,

$$\begin{aligned} \mathbb{E}^{\theta_\star} \left( \ln \frac{h(f^{\theta_\star}(Y_{-\infty:0}); Y_1)}{h(f^\theta(Y_{-\infty:0}); Y_1)} \right) \\ = \mathbb{E}^{\theta_\star} \left( \mathbb{E}^{\theta_\star} \left[ \ln \frac{h(f^{\theta_\star}(Y_{-\infty:0}); Y_1)}{h(f^\theta(Y_{-\infty:0}); Y_1)} \middle| Y_{-\infty:0} \right] \right). \end{aligned} \quad (34)$$

Under the stated assumptions,  $H(X_0; \cdot) = H(f^{\theta_\star}(Y_{-\infty:0}); \cdot)$  and (33) shows that  $H(f^{\theta_\star}(Y_{-\infty:0}); \cdot) = \mathbb{P}^{\theta_\star}[\cdot | Y_{-\infty:0}]$ . Therefore, the RHS of (34) is nonnegative as the expectation of a conditional Kullback-Leibler divergence. This shows that  $\theta_\star \in \Theta_\star = \operatorname{argmax}_{\theta \in \Theta} \mathbb{E}^{\theta_\star} (\ln h(f^\theta(Y_{-\infty:0}); Y_1))$ . Assume now that  $\theta \in \Theta_\star$ .



Then, according to (34),  $\mathbb{P}^{\theta_\star}$ -a.s., the probability measures  $H(f^{\theta_\star}\langle Y_{-\infty:0} \rangle; \cdot)$  and  $H(f^\theta\langle Y_{-\infty:0} \rangle; \cdot)$  are equal, so that under (a),

$$f^{\theta_\star}\langle Y_{-\infty:0} \rangle = f^\theta\langle Y_{-\infty:0} \rangle, \quad \mathbb{P}^{\theta_\star}\text{-a.s.} \quad (35)$$

Under (b), this implies that  $\theta = \theta_\star$ .  $\square$

### 2.3. Examples

#### 2.3.1. The Poisson threshold in misspecified models

Let  $K$  be a compact set of  $\mathbb{R}^5$  and let  $\Theta$  be the following (compact) set of parameters

$$\Theta = \{\theta = (\omega, a, b, c, d) \in K : \min(\omega, a, b, a + c, b + d) \geq \underline{\alpha}, a \vee (a + c) \leq \bar{\alpha} < 1\} . \quad (36)$$

where  $(\underline{\alpha}, \bar{\alpha}) \in (0, \infty) \times (0, 1)$ . Assume that the observations  $(Y_n)_{n \in \mathbb{Z}}$  are integer-valued and satisfy the following assumptions:

**(C1)**  $\{Y_n\}_{n \in \mathbb{Z}}$  is a strict-sense stationary and ergodic stochastic process

**(C2)**  $\mathbb{E}_\star[\ln(1 + Y_0)] < \infty$ .

The Poisson threshold autoregression model described in Example 3 may be rewritten as in (25), by setting  $\mathbf{X} = [\underline{\alpha}, \infty)$ ,  $\mathbf{Y} = \mathbb{N}$  and

$$f_y^\theta(x) = \omega + ax + by + (cx + dy) \mathbb{1}\{y \notin (L, U)\} , \quad (37)$$

$$h(x; y) = \frac{dH(x; \cdot)}{d\mu}(y) = \exp(-x)x^y/y! , \quad (38)$$

$$\theta = (\omega, a, b, c, d) ,$$

where  $\mu$  is the counting measure on  $\mathbb{N}$ . Note that  $f_y^\theta(x) = \omega + a^\theta(y)x + b^\theta(y)$  where  $a^\theta(y) = a + c \mathbb{1}\{y \notin (L, U)\}$  and  $b^\theta(y) = by + dy \mathbb{1}\{y \notin (L, U)\}$  so that for all  $(\theta, y) \in \Theta \times \mathbf{Y}$ ,  $|f_y^\theta(x) - f_y^\theta(x')| \leq \bar{\alpha}|x - x'|$ . Moreover, using (27), we have for all  $s \leq t$ ,

$$f^\theta\langle y_{s:t} \rangle(x) = x \prod_{\ell=s}^t a^\theta(y_\ell) + \sum_{j=0}^{t-s} [\omega + b^\theta(y_{t-j})] \prod_{\ell=1}^{j-1} a^\theta(y_{t-\ell}) . \quad (39)$$

With these definitions, let  $\theta_{n,x}$  be the Maximum Likelihood estimator associated to the likelihood function  $\mathbb{L}_{n,x}^\theta\langle Y_{1:n} \rangle$  as defined in (28) and (29).

**Theorem 25.** Assume (C1-2). Then, for all  $x \in \mathbf{X}$ ,  $\lim_{n \rightarrow \infty} d(\theta_{n,x}, \Theta_\star) = 0$ ,  $\mathbb{P}_{\theta_\star}$ -a.s. where

$$\Theta_\star := \operatorname{argmax}_{\theta \in \Theta} \mathbb{E}_\star(\bar{\ell}^\theta\langle Y_{-\infty:0} \rangle) , \quad (40)$$

where  $\bar{\ell}^\theta \langle Y_{-\infty:0} \rangle$ , defined in (30), can be written as:

$$\bar{\ell}^\theta \langle Y_{-\infty:0} \rangle = Y_0 \ln(f^\theta \langle Y_{-\infty:-1} \rangle) - f^\theta \langle Y_{-\infty:-1} \rangle - \ln Y_0! \quad (41)$$

$$f^\theta \langle Y_{-\infty:n} \rangle = \sum_{j=0}^{\infty} [\omega + b^\theta(Y_{n-j})] \prod_{\ell=1}^{j-1} a^\theta(Y_{n-\ell}), \quad \forall (n, \theta) \in \mathbb{Z} \times \Theta. \quad (42)$$

*Proof.* According to Theorem 21, it is sufficient to check (B2-3). (B2) clearly holds. Assumption (C2) allows to apply Lemma 22 so that (B3)-(ii) holds. Using Remark 20 shows that Assumption (B3)-(iii) is satisfied. It remains to check (B3)-(i). By (38), for all  $(x, x') \in [\underline{\alpha}, \infty)^2$  and  $y \in \Upsilon$ ,

$$\begin{aligned} |\ln h(x; y) - \ln h(x'; y)| &\leq \left( y \sup_{(x, x') \in [\underline{\alpha}, \infty)^2} \frac{|\ln(x) - \ln(x')|}{|x - x'|} + 1 \right) |x - x'| \\ &\leq (y/\underline{\alpha} + 1) |x - x'| \end{aligned}$$

Thus,

$$\begin{aligned} &\sup_{\theta \in \Theta} |\ln h(f^\theta \langle Y_{1:k-1} \rangle(x); Y_k) - \ln h(f^\theta \langle Y_{-\infty:k-1} \rangle; Y_k)| \\ &\leq (Y_k/\underline{\alpha} + 1) \sup_{\theta \in \Theta} |f^\theta \langle Y_{1:k-1} \rangle(x) - f^\theta \langle Y_{-\infty:k-1} \rangle| \\ &= (Y_k/\underline{\alpha} + 1) \sup_{\theta \in \Theta} \left| x \prod_{\ell=1}^{k-1} a^\theta(Y_\ell) + f^\theta \langle Y_{-\infty:0} \rangle \prod_{i=0}^{k-1} a^\theta(Y_i) \right| \\ &\leq (Y_k/\underline{\alpha} + 1) |x + \bar{\alpha} f^\theta \langle Y_{-\infty:0} \rangle| \bar{\alpha}^{k-1}, \end{aligned}$$

which converges to 0 as  $k$  goes to infinity by applying Lemma 36 under (C2).  $\square$

#### 2.4. The Poisson threshold in well-specified models

Let  $K$  be a compact set of  $\mathbb{R}^5$  and let  $\Theta$  be the following (compact) set of parameters

$$\begin{aligned} \Theta &= \{\theta = (\omega, a, b, c, d) \in K : \\ &\quad \min(\omega, a, b, a + c, b + d) \geq \underline{\alpha}, (a + b + c + d) \vee a \leq \bar{\alpha} < 1\}. \end{aligned} \quad (43)$$

where  $(\underline{\alpha}, \bar{\alpha}) \in (0, \infty) \times (0, 1)$ . We assume that  $(Y_k)$  is the observation process of Poisson threshold model as described in Example 3 with  $(\omega, a, b, c, d) = (\omega_\star, a_\star, b_\star, c_\star, d_\star) = \theta_\star$ .

**Proposition 26.** *Assume that  $\theta_\star \in \Theta$  and that  $(L, U) \cap \mathbb{N} \neq \emptyset$ . Then, for all  $x \in \mathbb{X}$ ,  $\lim_{n \rightarrow \infty} \theta_{n,x} = \theta_\star$ ,  $\mathbb{P}^{\theta_\star}$ -a.s..*

*Proof.* Let  $\{(X_k, Y_k), k \in \mathbb{Z}\}$  satisfying the recursions given by Example 3 with  $(\omega, a, b, c, d) = (\omega_\star, a_\star, b_\star, c_\star, d_\star) = \theta_\star$ . Proposition 14 shows that (C1-2) hold so that Theorem 25 applies. It thus remains to show that  $\Theta_\star = \{\theta_\star\}$ . This follows from Proposition 23 provided we show that

- (a)  $X_0 = f^{\theta_\star} \langle Y_{-\infty:0} \rangle$
- (b)  $H(x; \cdot) = H(x'; \cdot)$  implies that  $x = x'$ ,
- (c)  $f^{\theta_\star} \langle Y_{-\infty:0} \rangle = f^\theta \langle Y_{-\infty:0} \rangle$ ,  $\mathbb{P}^{\theta_\star}$ -a.s., implies that  $\theta = \theta_\star$ ,

Define

$$a_\star(y) = a_\star + c_\star \mathbb{1}\{y \notin (L, U)\}, \quad b_\star(y) = b_\star y + d_\star y \mathbb{1}\{y \notin (L, U)\},$$

First note that for all  $m \geq 0$ ,

$$X_0 = f^{\theta_\star} \langle Y_{-m:0} \rangle (X_{-m}) = X_{-m} \prod_{\ell=-m}^0 a_\star(Y_\ell) + \sum_{j=0}^m [\omega_\star + b_\star(Y_{-j})] \prod_{\ell=1}^j a_\star(Y_{-\ell}).$$

By applying Lemma 36 and Proposition 14, we have

$$X_{-m} \prod_{\ell=-m}^0 a_\star(Y_\ell) \leq \bar{\alpha}^{m+1} X_{-m} \xrightarrow{m \rightarrow \infty} 0, \quad \mathbb{P}^{\theta_\star}\text{-a.s.}$$

so that

$$X_0 = \lim_{m \rightarrow \infty} f^{\theta_\star} \langle Y_{-m:0} \rangle (X_{-m}) = f^{\theta_\star} \langle Y_{-\infty:0} \rangle, \quad \mathbb{P}^{\theta_\star}\text{-a.s.} \quad (44)$$

where  $f^{\theta_\star} \langle Y_{-\infty:0} \rangle$  is defined in (42). Thus, (a) holds. (b) also clearly holds since  $H(x, \cdot)$  is a Poisson distribution of parameter  $x$ . It remains to check (c). If  $\mathbb{P}^{\theta_\star}$ -a.s.,  $f^{\theta_\star} \langle Y_{-\infty:0} \rangle = f^\theta \langle Y_{-\infty:0} \rangle$ , then, by stationarity of the  $\{Y_n\}_{n \in \mathbb{Z}}$ , we have: for all  $t \in \mathbb{Z}$ ,  $X_t = X'_t$ ,  $\mathbb{P}^{\theta_\star}$ -a.s. where we set  $X'_t := f^\theta \langle Y_{-\infty:t} \rangle$ . This implies that  $X'_t = f_{Y_t}^\theta \circ f_{Y_{t-1}}^\theta (X'_{t-2}) = f_{Y_t}^\theta \circ f_{Y_{t-1}}^\theta (X_{t-2})$ ,  $\mathbb{P}^{\theta_\star}$ -a.s., so that,

$$\begin{aligned} & \omega + a^\theta(Y_t) [\omega + a^\theta(Y_{t-1})X_{t-2} + b^\theta(Y_{t-1})] + b^\theta(Y_t) \\ &= \omega_\star + a_\star(Y_t) [\omega_\star + a_\star(Y_{t-1})X_{t-2} + b_\star(Y_{t-1})] + b_\star(Y_t), \quad \mathbb{P}^{\theta_\star}\text{-a.s.} \end{aligned}$$

Since  $\mathbb{P}^{\theta_\star}[(Y_{t-1}, Y_t) = (k, \ell) | Y_{-\infty:t-2}] \neq 0$ , for all  $(k, \ell) \in \mathbb{N}^2$  and  $X_{t-2}$  is  $\sigma^\theta(Y_\ell, \ell \leq t-2)$ -measurable, we obtain that, for all  $(k, \ell) \in \mathbb{N}^2$ ,  $\mathbb{P}^{\theta_\star}$ -a.s.,

$$\begin{aligned} & \omega + a^\theta(k) [\omega + a^\theta(\ell)X_{t-2} + b^\theta(\ell)] + b^\theta(k) \\ &= \omega_\star + a_\star(k) [\omega_\star + a_\star(\ell)X_{t-2} + b_\star(\ell)] + b_\star(k) \end{aligned}$$

Fix  $\ell \in \mathbb{N}$ . Then, recalling that  $a^\theta(k)$  is bounded in  $k$  and that  $b^\theta(k) \sim_{k \rightarrow \infty} bk$ , we obtain that  $b = b_\star$ . Fix now  $k \in \mathbb{N}$  and take the equivalent of the previous equation as  $\ell$  goes to infinity, we then obtain  $a^\theta(k)b\ell = a_\star(k)b_\star\ell$  for all  $k \in \mathbb{N}$  which can also be written as

$$a + c \mathbb{1}_{k \notin (L, U)} = a_\star + c_\star \mathbb{1}_{k \notin (L, U)}$$

so that  $a = a_\star$  and  $c = c_\star$  by using that  $(L, U) \cap \mathbb{N} \neq \emptyset$ . Finally, using  $b = b_\star$ ,  $a = a_\star$  and  $c = c_\star$ , we have  $\mathbb{P}^{\theta_\star}$ -a.s. for all  $k \in \mathbb{N}$ ,

$$\omega + a_\star(k)X_{t-1} + b_\star k + dk \mathbb{1}_{k \notin (L, U)} = \omega_\star + a_\star(k)X_{t-1} + b_\star k + d_\star k \mathbb{1}_{k \notin (L, U)}$$

so that

$$\omega + dk \mathbb{1}_{k \notin (L, U)} = \omega_\star + d_\star k \mathbb{1}_{k \notin (L, U)} .$$

which again implies that  $\omega = \omega_\star$  and  $d = d_\star$  since  $(L, U) \cap \mathbb{N} \neq \emptyset$ .  $\square$

## 2.5. The Log-linear Poisson autoregression in misspecified Models

Let  $\Theta$  be the following (compact) set of parameters

$$\Theta = \left\{ \theta = (d, a, b) \in \mathbb{R}^3 : |d| \leq \tilde{d}, \quad |a| \leq \tilde{a} < 1, \quad |b| \leq \tilde{b} \right\} . \quad (45)$$

where  $\tilde{d}$ ,  $\tilde{a}$ ,  $\tilde{b}$  are positive constants of  $\mathbb{R}$ . Assume that the observations  $(Y_n)_{n \in \mathbb{Z}}$  are integer-valued and satisfy the assumptions **(C1-2)**. The Log-linear Poisson autoregression model described in (17) may be rewritten as in (25), by setting  $\mathbf{X} = \mathbb{R}$ ,  $\mathbf{Y} = \mathbb{N}$ ,  $\theta = (d, a, b)$ , and

$$f_y^\theta(x) = d + ax + b \ln(1 + y) , \quad (46)$$

$$h(x; y) = \frac{dH(x; \cdot)}{d\mu}(y) = \exp(-e^x) e^{xy} / y! , \quad (47)$$

where  $\mu$  is the counting measure on  $\mathbb{N}$ . Using (27), we have for all  $s \leq t$ ,

$$f^\theta \langle y_{s:t} \rangle(x) = d \frac{1 - a^{t-s+1}}{1 - a} + a^{t-s+1} x + b \sum_{j=0}^{t-s} a^j \ln(1 + y_{t-j}) . \quad (48)$$

With these definitions, let  $\theta_{n,x}$  be the Maximum Likelihood estimator associated to the likelihood function  $\mathbb{L}_{n,x}^\theta \langle Y_{1:n} \rangle$  as defined in (28) and (29).

**Theorem 27.** *Assume **(C1-2)**. Then, for all  $x \in \mathbf{X}$ ,  $\lim_{n \rightarrow \infty} d(\theta_{n,x}, \Theta_\star) = 0$ ,  $\mathbb{P}_{\theta_\star}$ -a.s., where*

$$\Theta_\star := \operatorname{argmax}_{\theta \in \Theta} \mathbb{E}_\star \left( Y_0 f^\theta \langle Y_{-\infty:-1} \rangle - e^{f^\theta \langle Y_{-\infty:-1} \rangle} - \ln Y_0! \right) , \quad (49)$$

$$f^\theta \langle Y_{-\infty:n} \rangle := \frac{d}{1-a} + b \sum_{j=0}^{\infty} a^j \ln(1 + Y_{n-j}) , \quad \forall (n, \theta) \in \mathbb{Z} \times \Theta . \quad (50)$$

*Proof.* According to Theorem 21, it is sufficient to check **(B2-3)**. **(B2)** clearly holds. Using Remark 20, since  $\mathbf{Y} = \mathbb{N}$ , we only need to check **(B3)**-(i) and **(B3)**-(ii). First note that

$$\sup_{\theta \in \Theta} |f^\theta \langle Y_{-\infty:0} \rangle| \leq \tilde{d}/(1 - \tilde{a}) + \tilde{b} \sum_{j=0}^{\infty} \tilde{a}^j \ln(1 + Y_{-j}) < \infty, \quad \mathbb{P}_\star\text{-a.s.} \quad (51)$$

which is finite according to **(C2)** by using Lemma 36. Now, write for all  $\theta =$

$(d, a, b) \in \Theta$ ,

$$\begin{aligned} & |f^\theta \langle Y_{-m:0} \rangle(x) - f^\theta \langle Y_{-\infty:0} \rangle| \\ &= |a|^{m+1} \left| -\frac{d}{1-a} + x + b \sum_{\ell=0}^{\infty} a^\ell \ln(1 + Y_{-m-1-\ell}) \right| \\ &\leq |\tilde{a}|^{m+1} \left( \frac{\tilde{d}}{1-\tilde{a}} + |x| + \tilde{b} \sum_{\ell=0}^{\infty} \tilde{a}^\ell \ln(1 + Y_{-m-1-\ell}) \right). \end{aligned}$$

By (C2) and by applying Lemma 36, the right-hand side (which does not depend on  $\theta$ ) converges to 0 as  $m$  goes to infinity. Thus, (i) holds. We now turn to (ii).

$$\begin{aligned} & \sup_{\theta \in \Theta} |\ln h(f^\theta \langle Y_{1:k-1} \rangle(x); Y_k) - \bar{\ell}^\theta \langle Y_{-\infty:k} \rangle| \leq \\ & Y_k \sup_{\theta \in \Theta} |f^\theta \langle Y_{1:k-1} \rangle(x) - f^\theta \langle Y_{-\infty:k-1} \rangle| + \sup_{\theta \in \Theta} \left| e^{f^\theta \langle Y_{1:k-1} \rangle(x)} - e^{f^\theta \langle Y_{-\infty:k-1} \rangle} \right|. \end{aligned} \quad (52)$$

Consider the first term in the rhs. It follows immediately from (48) and (50) that

$$f^\theta \langle Y_{1:k-1} \rangle(x) - f^\theta \langle Y_{-\infty:k-1} \rangle = a^{k-1} (x - f^\theta \langle Y_{-\infty:0} \rangle). \quad (53)$$

This implies that, for all  $k \geq 1$ ,

$$Y_k \sup_{\theta \in \Theta} |f^\theta \langle Y_{1:k-1} \rangle(x) - f^\theta \langle Y_{-\infty:k-1} \rangle| \leq Y_k \tilde{a}^{k-1} (x + \sup_{\theta \in \Theta} |f^\theta \langle Y_{-\infty:0} \rangle|),$$

which converges  $\mathbb{P}$ -a.s. to 0 as  $k$  goes to infinity according to (51) and by applying Lemma 36 under (C2). Moreover, (53) also implies that

$$\begin{aligned} & \left| e^{f^\theta \langle Y_{0:k-1} \rangle(x)} - e^{f^\theta \langle Y_{-\infty:k-1} \rangle} \right| = e^{f^\theta \langle Y_{-\infty:k-1} \rangle} \left| e^{a^{k-1} (x - f^\theta \langle Y_{-\infty:0} \rangle)} - 1 \right| \\ & \leq |a|^{k-1} |x - f^\theta \langle Y_{-\infty:0} \rangle| e^{|x - f^\theta \langle Y_{-\infty:0} \rangle| + f^\theta \langle Y_{-\infty:k-1} \rangle}, \end{aligned}$$

so that the second term of the rhs of (52) is bounded according to

$$\begin{aligned} & \sup_{\theta \in \Theta} \left| e^{f^\theta \langle Y_{0:k-1} \rangle(x)} - e^{f^\theta \langle Y_{-\infty:k-1} \rangle} \right| \\ & \leq \sup_{\theta \in \Theta} \left( |x - f^\theta \langle Y_{-\infty:0} \rangle| e^{|x - f^\theta \langle Y_{-\infty:0} \rangle|} \right) \times \tilde{a}^{k-1} \sup_{\theta \in \Theta} e^{|f^\theta \langle Y_{-\infty:k-1} \rangle|}. \end{aligned}$$

To complete the proof, it is thus sufficient to show that

$$\lim_{k \rightarrow \infty} \tilde{a}^k \exp \left\{ \sup_{\theta \in \Theta} |f^\theta \langle Y_{-\infty:k-1} \rangle| \right\} = 0, \quad \mathbb{P}_{\theta_\star}\text{-a.s.}$$

But this is straightforward by applying Lemma 36 since by (51) and by setting  $V_k := \exp \{ \sup_{\theta \in \Theta} |f^\theta \langle Y_{-\infty:k-1} \rangle| \}$ , we have

$$\mathbb{E}_\star [(\ln V_1)_+] \leq \frac{\tilde{d}}{1-\tilde{a}} + \tilde{b} \sum_{j=0}^{\infty} \tilde{a}^j \mathbb{E}_\star [\ln(1 + Y_{-j})] = \frac{\tilde{d} + \tilde{b} \mathbb{E}_\star [\ln(1 + Y_0)]}{1-\tilde{a}},$$

which is finite by (C2).  $\square$

## 2.6. Log-linear Poisson autoregression in well-specified models

Let  $\Theta$  be the following (compact) set of parameters

$$\Theta = \{(d, a, b) \in \mathbb{R}^3; |d| \leq \tilde{d}, |a + b| \vee |a| \vee |b| \leq \tilde{\alpha} < 1\}. \quad (54)$$

where  $(\tilde{d}, \tilde{\alpha}) \in \mathbb{R}_+^+ \times (0, 1)$ . We assume that  $\{Y_n\}_{n \in \mathbb{Z}}$  is the observation process of the Log-linear Poisson autoregression model described in (17) with parameters  $(d, a, b) = (d_*, a_*, b_*) = \theta_*$ .

**Proposition 28.** *Assume that  $\theta_* \in \Theta$ . Then, for all  $x \in \mathbb{X}$ ,*

$$\lim_{n \rightarrow \infty} \theta_{n,x} = \theta_*, \quad \mathbb{P}^{\theta_*}\text{-a.s.}$$

*Proof.* Let  $\{(X_k, Y_k), k \in \mathbb{Z}\}$  satisfying (17) with  $(d, a, b) = (d_*, a_*, b_*) = \theta_*$ . Proposition 17 shows that (C1-2) hold so that Theorem 27 applies. It thus remains to show that  $\Theta_* = \{\theta_*\}$ . This follows from Proposition 23 provided we show that

- (a)  $X_0 = f^{\theta_*} \langle Y_{-\infty:0} \rangle$ ,  $\mathbb{P}^{\theta_*}\text{-a.s.}$ ,
- (b)  $H(x; \cdot) = H(x'; \cdot)$  implies that  $x = x'$ ,
- (c)  $f^{\theta_*} \langle Y_{-\infty:0} \rangle = f^\theta \langle Y_{-\infty:0} \rangle$ ,  $\mathbb{P}^{\theta_*}\text{-a.s.}$ , implies that  $\theta = \theta_*$ ,

First note that for all  $m \geq 0$ ,

$$X_0 = f^{\theta_*} \langle Y_{-m:0} \rangle (X_{-m}) = d_* \frac{1 - a_*^{m+1}}{1 - a_*} + a_*^{m+1} X_{-m} + b_* \sum_{j=0}^m a_*^j \ln(1 + Y_{-j-1}).$$

By applying Lemma 36 and Proposition 17, we have  $\lim_{m \rightarrow \infty} a_*^{m+1} X_{-m} = 0$ ,  $\mathbb{P}^{\theta_*}\text{-a.s.}$ , so that

$$X_0 = \lim_{m \rightarrow \infty} f^{\theta_*} \langle Y_{-m:0} \rangle (X_{-m}) = f^{\theta_*} \langle Y_{-\infty:0} \rangle, \quad \mathbb{P}^{\theta_*}\text{-a.s.} \quad (55)$$

where  $f^{\theta_*} \langle Y_{-\infty:0} \rangle$  is defined in (50). Thus, (a) holds. (b) also clearly holds since  $H(x, \cdot)$  is a Poisson distribution of parameter  $e^x$ . It remains to check (c). If  $\mathbb{P}^{\theta_*}\text{-a.s.}$ ,  $f^{\theta_*} \langle Y_{-\infty:0} \rangle = f^\theta \langle Y_{-\infty:0} \rangle$ , then, by definition of  $f^\theta \langle Y_{-\infty:0} \rangle$ ,

$$\frac{d_*}{1 - a_*} - \frac{d}{1 - a} + \sum_{j=0}^{\infty} (b_* a_*^j - b a^j) \ln(1 + Y_{-j}) = 0, \quad \mathbb{P}^{\theta_*}\text{-a.s.}$$

Conditionally on  $\sigma(Y_m; m \leq -1)$ ,  $Y_0$  is a Poisson random variable with a positive intensity; thus, the lhs is constant only if  $b_* = b$ . This implies that

$$\frac{d_*}{1 - a_*} - \frac{d}{1 - a} + b \sum_{j=1}^{\infty} (a_*^j - a^j) \ln(1 + Y_{-j}) = 0, \quad \mathbb{P}^{\theta_*}\text{-a.s.}$$

By the same argument, the lhs is constant conditionally on  $\sigma(Y_m; m \leq -2)$  only if  $a_* = a$ . In that case, the previous equality writes:  $d_* - d = 0$  which completes the proof.  $\square$

### 3. Proofs of Theorem 8, Lemma 9 and Lemma 11

The proof roughly follows the lines of Henderson et al. (2011) with the difference that we relax the Lipschitz assumption and introduce a drift function  $W$ . In all this section,  $(X, d)$  is a Polish (complete, separable and metric) space and denote by  $\mathcal{X}$  its associated Borel  $\sigma$ -field. A *totally separating system of metrics*  $\{d_n, n \in \mathbb{N}\}$  for  $X$  is a set of metrics such that for all fixed  $x, x' \in X$ , the sequence  $\{d_n(x, x'), n \in \mathbb{N}\}$  is nondecreasing in  $n$  and  $\lim_{n \rightarrow \infty} d_n(x, x') = \mathbb{1}_{x \neq x'}$ . A metric  $d$  on  $X$  induces a Wasserstein distance between probability measures  $\mu_1$  and  $\mu_2$  on  $(X, \mathcal{X})$  defined by:

$$\|\mu_1 - \mu_2\|_d = \inf \left\{ \int \mu(dx, dx') d(x, x') : \mu \in \mathcal{M}(\mu_1, \mu_2) \right\}, \quad (56)$$

where  $\mathcal{M}(\mu_1, \mu_2)$  is the set of probability measures  $\mu$  on  $(X^2, \mathcal{X}^{\otimes 2})$  such that

$$\mu(A \times X) = \mu_1(A), \quad \mu(X \times A) = \mu_2(A), \quad \text{for all } A \in \mathcal{X}.$$

Since  $X$  is a separable metric space, the Kantorowich-Rubinstein duality theorem applies (see for example (Dudley, 2002, Theorem 11.8.2)) and we have

$$\|\mu_1 - \mu_2\|_d = \sup \{ \mu_1(f) - \mu_2(f) : \text{Lip}(f; d) \leq 1 \}, \quad (57)$$

where

$$\text{Lip}(f; d) = \sup \left\{ \frac{|f(x) - f(x')|}{d(x, x')} : x, x' \in X, x \neq x' \right\}.$$

Recall the definition of an asymptotically strong Feller kernel, first introduced by Hairer and Mattingly (2006):

**Definition 29.** A Markov kernel  $Q$  is asymptotically strong Feller if, for all  $x \in X$ , there exist a totally separating system of metrics  $\{d_n, n \in \mathbb{N}\}$  for  $X$  and a sequence of integers  $\{t_n, n \in \mathbb{N}\}$  such that

$$\lim_{\gamma \rightarrow 0} \limsup_{n \rightarrow \infty} \sup_{x' \in B(x, \gamma)} \|Q^{t_n}(x, \cdot) - Q^{t_n}(x', \cdot)\|_{d_n} = 0.$$

where  $B(x, \gamma)$  is the open ball of radius  $\gamma$  with respect to  $d$  and centered at  $x$ .

The following theorem is taken from Hairer and Mattingly (2006) and provide conditions for obtaining uniqueness of the invariant probability measure.

**Theorem 30.** Assume that the Markov kernel  $Q$  is asymptotically strong Feller and admits a reachable point  $x \in X$ . Then,  $Q$  has at most one stationary distribution.

*Proof of Theorem 8.* Under (A1), (Tweedie, 1988, Theorem 2) show that  $Q$  admits at least one stationary distribution. Since by (A2)  $Q$  admits a reachable point, we conclude by applying Theorem 30 provided that we can prove  $Q$  is asymptotically strong Feller. Denote

$$T = \inf \{i \in \mathbb{N} : U_i = 0\} \quad (58)$$

with the convention  $\inf \emptyset = \infty$ . We preface the proof by the following technical lemma:

**Lemma 31.** Let  $\{(X_k, X'_k, U_k), k \in \mathbb{N}\}$  be a Markov chain on  $(\mathbf{X}^2 \times \{0, 1\})$  with Markov kernel  $\bar{Q}$ , introduced in (A3). Then, for all real-valued nonnegative measurable function  $\varphi$  on  $\mathbf{X}^2$ ,  $n \in \mathbb{N}^*$ ,  $x, x' \in \mathbf{X}$  and  $u \in [0, 1]$ ,

$$\mathbb{E}_{\delta_x \otimes \delta_{x'} \otimes \mathcal{B}(u)}^{\bar{Q}} [\varphi(X_n^\#) \mathbb{1}_{\{T > n\}}] = \mathbb{E}_{\delta_x \otimes \delta_{x'}}^{Q^\#} \left[ \varphi(X_n^\#) \prod_{i=0}^{n-1} \alpha(X_i^\#) \right], \quad (59)$$

where  $Q^\#$  and  $\alpha$  are introduced in (A3),  $X_i^\# := (X_i, X'_i)$  and  $\mathcal{B}(u)$  is the Bernoulli distribution with parameter  $u$ .

*Proof.* The proof is by induction. Note first that (59) obviously holds for  $n = 1$ . Now, assume that (59) holds for some  $n \geq 1$ . Then, noting that  $\mathbb{1}_{\{T > n+1\}} = \prod_{i=1}^{n+1} U_i$  (where  $T$  is defined in (58))

$$\begin{aligned} \mathbb{P}_{\delta_x \otimes \delta_{x'} \otimes \mathcal{B}(u)}^{\bar{Q}} [\varphi(X_{n+1}^\#) \mathbb{1}_{\{T > n+1\}}] &= \mathbb{E}_{\delta_x \otimes \delta_{x'} \otimes \mathcal{B}(u)}^{\bar{Q}} \left[ \varphi(X_{n+1}^\#) \prod_{i=1}^{n+1} U_i \right] \\ &= \mathbb{E}_{\delta_x \otimes \delta_{x'} \otimes \mathcal{B}(u)}^{\bar{Q}} \left[ \mathbb{E}^{\bar{Q}}(U_{n+1} \varphi(X_{n+1}^\#) | X_n^\#, U_n) \prod_{i=1}^n U_i \right] \\ &= \mathbb{E}_{\delta_x \otimes \delta_{x'} \otimes \mathcal{B}(u)}^{\bar{Q}} \left[ \alpha(X_n^\#) Q^\# \varphi(X_n^\#) \prod_{i=1}^n U_i \right]. \end{aligned}$$

where the last equality follows from (5). Applying the induction assumption to the right-hand side of the inequality, we obtain

$$\begin{aligned} \mathbb{P}_{\delta_x \otimes \delta_{x'} \otimes \mathcal{B}(u)}^{\bar{Q}} [\varphi(X_{n+1}^\#) \mathbb{1}_{\{T > n+1\}}] &= \mathbb{E}_{\delta_x \otimes \delta_{x'}}^{Q^\#} \left[ \alpha(X_n^\#) Q^\# \varphi(X_n^\#) \prod_{i=0}^{n-1} \alpha(X_i^\#) \right] \\ &= \mathbb{E}_{\delta_x \otimes \delta_{x'}}^{Q^\#} \left[ \alpha(X_n^\#) \mathbb{E}^{Q^\#}(\varphi(X_{n+1}^\#) | X_n^\#) \prod_{i=0}^{n-1} \alpha(X_i^\#) \right] \\ &= \mathbb{E}_{\delta_x \otimes \delta_{x'}}^{Q^\#} \left[ \varphi(X_{n+1}^\#) \prod_{i=0}^n \alpha(X_i^\#) \right]. \end{aligned}$$

The proof is completed.  $\square$

Now, consider  $d_n(x, x') = 1 \wedge [nd(x, x')]$ . Obviously, for all fixed  $x, x' \in \mathbf{X}$ , the sequence  $\{d_n(x, x'), n \in \mathbb{N}\}$  is nondecreasing and  $\lim_{n \rightarrow \infty} d_n(x, x') = \mathbb{1}_{\{x \neq x'\}}$  so that  $\{d_n, n \in \mathbb{N}\}$  is a totally separating system of metrics. Moreover, the Kantorovich-Rubinstein duality theorem (56), (57) and the marginal conditions (3) yield: for all  $(x, x', u) \in \mathbf{X} \times \mathbf{B}(x, \gamma_x) \times [0, 1]$ ,

$$\begin{aligned} \|\delta_x Q^n - \delta_{x'} Q^n\|_{d_n} &\leq \mathbb{E}_{\delta_x \otimes \delta_{x'} \otimes \mathcal{B}(u)}^{\bar{Q}}(d_n(X_n, X'_n)) \\ &\leq \mathbb{P}_{\delta_x \otimes \delta_{x'} \otimes \mathcal{B}(u)}^{\bar{Q}}(T \leq n) + \mathbb{E}_{\delta_x \otimes \delta_{x'} \otimes \mathcal{B}(u)}^{\bar{Q}}(d_n(X_n, X'_n) \mathbb{1}_{\{T > n\}}), \quad (60) \end{aligned}$$



where we have used that  $d_n(x, y) \leq 1$ . First consider the second term of the right-hand side. Applying Lemma 31, combined with  $d_n(x, y) \leq nd(x, y)$ ,  $\alpha(x, x') \leq 1$  for all  $(x, x') \in \mathbf{X}^2$  and (8), yields

$$\begin{aligned} \mathbb{E}_{\delta_x \otimes \delta_{x'} \otimes \mathcal{B}(u)}^{\bar{Q}}(d_n(X_n, X'_n) \mathbb{1}\{T > n\}) &\leq n \mathbb{E}_{\delta_x \otimes \delta_{x'}}^{Q^\#} \left[ d(X_n, X'_n) \prod_{i=0}^{n-1} \alpha(X_i, X'_i) \right] \\ &\leq nD\rho^n d(x, x') . \end{aligned} \quad (61)$$

We now turn to the first term of the right-hand side of (60). By (7), we get

$$\begin{aligned} \mathbb{P}_{\delta_x \otimes \delta_{x'} \otimes \mathcal{B}(u)}^{\bar{Q}}(T \leq n) &= \sum_{k=0}^{n-1} \mathbb{P}_{\delta_x \otimes \delta_{x'} \otimes \mathcal{B}(u)}^{\bar{Q}}(T > k, U_{k+1} = 0) \\ &\leq \sum_{k=0}^{n-1} \mathbb{E}_{\delta_x \otimes \delta_{x'} \otimes \mathcal{B}(u)}^{\bar{Q}}[\mathbb{1}\{T > k\} d(X_k, X'_k) W(X_k, X'_k)] \end{aligned}$$

Applying Lemma 31 to the right-hand side, combined with  $\alpha \leq 1$  and (9), we obtain

$$\begin{aligned} \mathbb{P}_{\delta_x \otimes \delta_{x'} \otimes \mathcal{B}(u)}^{\bar{Q}}(T \leq n) &\leq \sum_{k=0}^{n-1} \mathbb{E}_{\delta_x \otimes \delta_{x'}}^{Q^\#} \left[ d(X_k, X'_k) W(X_k, X'_k) \prod_{i=0}^{k-1} \alpha(X_i, X'_i) \right] \\ &\leq Dd^{\zeta_1}(x, x') W^{\zeta_2}(x, x') \sum_{k=0}^{n-1} \rho^k \leq \frac{Dd^{\zeta_1}(x, x') W^{\zeta_2}(x, x')}{1 - \rho} . \end{aligned}$$

Plugging this and (61) into (60) yields: for all  $x \in \mathbf{X}$  and all  $x' \in \mathcal{B}(x, \gamma)$  where  $\gamma < \gamma_x$ ,

$$\|\delta_x Q^n - \delta_{x'} Q^n\|_{d_n} \leq D \left( n\rho^n \gamma + \gamma^{\zeta_1} \sup_{y \in \mathcal{B}(x, \gamma_x)} W^{\zeta_2}(x, y) / (1 - \rho) \right) ,$$

where  $\gamma_x$  is defined in (10). Thus, for all  $x \in \mathbf{X}$ ,

$$\lim_{\gamma \rightarrow 0} \limsup_{n \rightarrow \infty} \sup_{x' \in \mathcal{B}(x, \gamma)} \|Q^n(x, \cdot) - Q^n(x', \cdot)\|_{d_n} = 0 .$$

The proof is completed.  $\square$

*Proof of Lemma 9.* Since for all  $M > 0$ , the function  $x \mapsto x \wedge M$  is concave, we have for all  $n \in \mathbb{N}$ ,

$$Q^n(V \wedge M) \leq (Q^n V) \wedge M \leq [\lambda^n V + b/(1 - \lambda)] \wedge M .$$

By integrating with respect to  $\pi$ , we obtain that

$$\pi(V \wedge M) = \pi Q^n(V \wedge M) \leq \pi \{[\lambda^n V + \beta/(1 - \lambda)] \wedge M\} .$$

The Lebesgue convergence theorem yields by letting  $n$  goes to infinity

$$\pi(V \wedge M) \leq \beta/(1 - \lambda) \wedge M .$$

The proof follows by letting  $M$  goes to infinity.  $\square$

## Appendix A. Ergodicity of one-sided and two-sided sequences

Let  $(X, \mathcal{X})$  be a measurable space. Denote by  $S : X^{\mathbb{N}} \rightarrow X^{\mathbb{N}}$  and  $\tilde{S} : X^{\mathbb{Z}} \rightarrow X^{\mathbb{Z}}$  the shift operators defined by: for all  $\mathbf{x} = (x_t)_{t \in \mathbb{N}} \in X^{\mathbb{N}}$  and all  $\tilde{\mathbf{x}} = (\tilde{x}_t)_{t \in \mathbb{Z}} \in X^{\mathbb{Z}}$ ,

$$S(\mathbf{x}) = (y_t)_{t \in \mathbb{N}}, \quad \text{where } y_t = x_{t+1}, \quad \forall t \in \mathbb{N}, \quad (\text{A.1})$$

$$\tilde{S}(\tilde{\mathbf{x}}) = (\tilde{y}_t)_{t \in \mathbb{Z}}, \quad \text{where } \tilde{y}_t = \tilde{x}_{t+1}, \quad \forall t \in \mathbb{Z}. \quad (\text{A.2})$$

Note that  $\tilde{S}$  is invertible while  $S$  is not. Let  $P$  be a Markov kernel on  $(X, \mathcal{X})$ . Denote by  $\mathbb{P}_\mu$  the probability induced on  $(X^{\mathbb{N}}, \mathcal{X}^{\otimes \mathbb{N}})$  by a Markov chain of initial distribution  $\mu$  and Markov kernel  $P$  and write  $\mathbb{E}_\mu$  the associated expectation operator. If  $\mu = \pi$  is an invariant distribution for  $P$ , we can define a probability  $\tilde{\mathbb{P}}_\pi$  induced on  $(X^{\mathbb{Z}}, \mathcal{X}^{\otimes \mathbb{Z}})$  by the Markov kernel  $P$  and initial distribution  $\pi$ . Similarly, we write  $\tilde{\mathbb{E}}_\pi$  the associated expectation operator. Moreover,  $\tilde{\mathbb{P}}_\pi$  extends  $\mathbb{P}_\pi$  on  $\mathbb{Z}$  in the sense that for all  $A \in \mathcal{X}^{\otimes \mathbb{N}}$ ,  $\mathbb{P}_\pi(A) = \tilde{\mathbb{P}}_\pi(X^{\mathbb{Z}^+} \times A)$ , which can also be written as

$$\mathbb{P}_\pi = \tilde{\mathbb{P}}_\pi \circ p^{-1}, \quad (\text{A.3})$$

where  $p$  is the mapping from  $(X^{\mathbb{Z}}, \mathcal{X}^{\otimes \mathbb{Z}})$  to  $(X^{\mathbb{N}}, \mathcal{X}^{\otimes \mathbb{N}})$  defined by

$$p(\omega) = (\omega_n)_{n \in \mathbb{N}} \quad \text{where } \omega = (\omega_n)_{n \in \mathbb{Z}}. \quad (\text{A.4})$$

Define for all  $k \in \mathbb{N}$ ,  $X_k : X^{\mathbb{N}} \rightarrow X$  by

$$X_k(\omega) = \omega_k, \quad \text{where } \omega = (\omega_\ell)_{\ell \in \mathbb{N}} \in X^{\mathbb{N}},$$

and similarly, define for all  $k \in \mathbb{Z}$ ,  $\tilde{X}_k : X^{\mathbb{Z}} \rightarrow X$  by

$$\tilde{X}_k(\tilde{\omega}) = \tilde{\omega}_k, \quad \text{where } \tilde{\omega} = (\omega_\ell)_{\ell \in \mathbb{Z}} \in X^{\mathbb{Z}}. \quad (\text{A.5})$$

Recall that  $(\Omega, \mathcal{F}, \mathbb{P}, \tau)$  is a *measure-preserving* dynamical system if  $(\Omega, \mathcal{F}, \mathbb{P})$  is a probability space and  $\tau : \Omega \rightarrow \Omega$  is measurable such that  $\mathbb{P} \circ \tau^{-1} = \mathbb{P}$ . Moreover, a measure-preserving dynamical system  $(\Omega, \mathcal{F}, \mathbb{P}, \tau)$  is said to be *ergodic* if for all invariant subset  $A \in \mathcal{F}$ , i.e.  $\mathbb{1}_A = \mathbb{1}_A \circ S$ , we have  $\mathbb{P}(A) = 0$  or  $1$ . Recall that if  $\mathbb{1}_B = \mathbb{1}_A \circ S$ ,  $\mathbb{P}$ -a.s., then, there exists an invariant set  $A$  such that  $\mathbb{1}_A = \mathbb{1}_B$ ,  $\mathbb{P}$ -a.s. In the following,  $\tau^k : \Omega \rightarrow \Omega$  is the mapping  $\tau$  iterated  $k$  times, that is  $\tau^k = \tau \circ \dots \circ \tau$  and by convention  $\tau^0(\omega) = \omega$  for all  $\omega \in \Omega$ .

**Theorem 32.** *Assume that the Markov kernel  $P$  has a unique stationary distribution  $\pi$ . Then, the dynamical system  $(X^{\mathbb{N}}, \mathcal{X}^{\otimes \mathbb{N}}, \mathbb{P}_\pi, S)$  is ergodic.*

*Proof.* Let  $A \in \mathcal{X}^{\otimes \mathbb{N}}$  be an invariant set for  $(\mathbf{X}^{\mathbb{N}}, \mathcal{X}^{\otimes \mathbb{N}}, \mathbb{P}_\pi, S)$ , that is:  $\mathbb{1}_A = \mathbb{1}_A \circ S$ . We will show that  $\mathbb{P}_\pi(A) = 0$  or 1 by contradiction. Assume indeed that  $\mathbb{P}_\pi(A) \in (0, 1)$ . Using the Markov property and the fact that  $A$  is invariant,

$$\mathbb{E}_{X_k}(\mathbb{1}_A) = \mathbb{E}_\pi \left[ \mathbb{1}_A \circ S^k \middle| \mathcal{F}_k \right] = \mathbb{E}_\pi [\mathbb{1}_A | \mathcal{F}_k], \quad \mathbb{P}_\pi\text{-a.s.}$$

where  $\mathcal{F}_k = \sigma(X_0, \dots, X_k)$ . Therefore,  $\{(\mathbb{E}_{X_k}(\mathbb{1}_A), \mathcal{F}_k), k \in \mathbb{N}\}$  is a uniformly integrable martingale. By (Hall and Heyde, 1980, Corollary 2.2),  $\lim_{k \rightarrow \infty} \mathbb{E}_{X_k}(\mathbb{1}_A) = \mathbb{1}_A$ ,  $\mathbb{P}_\pi$ -a.s. and  $\lim_{k \rightarrow \infty} \mathbb{E}_\pi |\mathbb{E}_{X_k}(\mathbb{1}_A) - \mathbb{1}_A| = 0$ . Then,

$$\begin{aligned} \mathbb{E}_\pi(|\mathbb{1}_A - \mathbb{E}_{X_0}(\mathbb{1}_A)|) &= \mathbb{E}_\pi(|\mathbb{1}_A - \mathbb{E}_{X_0}(\mathbb{1}_A)| \circ S^k) \\ &= \mathbb{E}_\pi(|\mathbb{1}_A - \mathbb{E}_{X_k}(\mathbb{1}_A)|) = \lim_{k \rightarrow \infty} \mathbb{E}_\pi(|\mathbb{1}_A - \mathbb{E}_{X_k}(\mathbb{1}_A)|) = 0. \end{aligned}$$

so that  $\mathbb{1}_A = \mathbb{P}_{X_0}(A)$ ,  $\mathbb{P}_\pi$ -a.s. Setting

$$\Gamma_A := \{x \in \mathbf{X}, \mathbb{P}_x(A) = 1\}, \quad (\text{A.6})$$

we then obtain  $\mathbb{1}_A = \mathbb{1}_{\Gamma_A}(X_0)$ ,  $\mathbb{P}_\pi$ -a.s. Combining it with the fact that  $A$  is invariant, we get for all  $k \in \mathbb{N}$ ,

$$\mathbb{1}_A = \mathbb{1}_A \circ S^k = \mathbb{1}_{\Gamma_A}(X_0 \circ S^k) = \mathbb{1}_{\Gamma_A}(X_k), \quad \mathbb{P}_\pi\text{-a.s.} \quad (\text{A.7})$$

Now, let  $\pi_A(\cdot) = \alpha^{-1} \pi(\Gamma_A \cap \cdot)$  where  $\alpha = \mathbb{P}_\pi(A) \neq 0$ . By definition of  $\pi_A$  and by using (A.7) with  $k = 0$  and  $k = 1$ , we get for all  $B \in \mathcal{X}$ ,

$$\begin{aligned} \mathbb{P}_{\pi_A}(X_1 \in B) &= \alpha^{-1} \mathbb{P}_\pi(\{X_1 \in B\} \cap \{X_0 \in \Gamma_A\}) \\ &= \alpha^{-1} \mathbb{P}_\pi(\{X_1 \in B\} \cap \{X_1 \in \Gamma_A\}) \\ &= \alpha^{-1} \mathbb{P}_\pi(X_1 \in B \cap \Gamma_A) = \alpha^{-1} \pi(B \cap \Gamma_A) = \pi_A(B), \end{aligned}$$

showing that  $\pi_A$  is a stationary distribution for the Markov kernel  $Q$ . Since  $A$  is an invariant set,  $A^c$  is also an invariant set and thus,  $\pi_{A^c}$  is also a stationary distribution for the Markov kernel  $Q$ . Since by assumption there exists a unique stationary distribution, we have that  $\pi_A = \pi_{A^c}$  which is not possible since these probability measures have disjoint supports (indeed by (A.6), we have  $\Gamma_A \cap \Gamma_{A^c} = \emptyset$ ).  $\square$

**Theorem 33.** *Assume that the dynamical system  $(\mathbf{X}^{\mathbb{N}}, \mathcal{X}^{\otimes \mathbb{N}}, \mathbb{P}_\pi, S)$  is ergodic. Then, the dynamical system  $(\mathbf{X}^{\mathbb{Z}}, \mathcal{X}^{\otimes \mathbb{Z}}, \tilde{\mathbb{P}}_\pi, \tilde{S})$  is ergodic.*

*Proof.* Let  $A$  be an invariant set for the dynamical system  $(\mathbf{X}^{\mathbb{Z}}, \mathcal{X}^{\otimes \mathbb{Z}}, \tilde{\mathbb{P}}_\pi, \tilde{S})$ , that is  $\mathbb{1}_A = \mathbb{1}_A \circ \tilde{S}$ . We now show that  $\tilde{\mathbb{P}}_\pi(A) = 0$  or 1.

Note first that  $\mathcal{X}^{\otimes \mathbb{Z}} = \sigma(\mathcal{F}_{-k}, k \in \mathbb{N})$  where  $\mathcal{F}_\ell = \sigma(\tilde{X}_i, \ell \leq i < \infty)$  and  $\tilde{X}_i$  is defined in (A.5). This allows to apply the approximation Lemma (see for example (Gray, 2009, Corollary 1.5.3)) showing that for all  $\epsilon > 0$ , there exists  $k_\epsilon \in \mathbb{N}$  and a  $\mathcal{F}_{-k_\epsilon}$ -measurable random variable  $Z_\epsilon$  such that  $\tilde{\mathbb{E}}_\pi(|Z_\epsilon|) < \infty$

and  $\tilde{\mathbb{E}}_\pi|\mathbb{1}_A - Z_\epsilon| \leq \epsilon$ . Then, setting  $Y_\epsilon = Z_\epsilon \circ \tilde{S}^{k_\epsilon} \in \mathcal{F}_0$  and using that  $A$  is an invariant set, we obtain

$$\tilde{\mathbb{E}}_\pi|\mathbb{1}_A - Y_\epsilon| = \tilde{\mathbb{E}}_\pi|\mathbb{1}_A \circ \tilde{S}^k - Z_\epsilon \circ \tilde{S}^k| = \tilde{\mathbb{E}}_\pi|\mathbb{1}_A - Z_\epsilon| \leq \epsilon .$$

The positive real number  $\epsilon$  being arbitrary, there exists  $Y$  such that  $\tilde{\mathbb{E}}_\pi|Y| < \infty$  and  $\mathbb{1}_A = Y$ ,  $\tilde{\mathbb{P}}_\pi$ -a.s. which implies that  $1 = \tilde{\mathbb{P}}_\pi(\mathbb{1}_A = Y) \leq \tilde{\mathbb{P}}_\pi(Y \in \{0, 1\}) \leq 1$ . Thus, there exists  $B \in \mathcal{F}_0$  such that

$$\mathbb{1}_B = Y = \mathbb{1}_A , \quad \tilde{\mathbb{P}}_\pi\text{-a.s.} \quad (\text{A.8})$$

Eq. (A.8) and the invariance of  $A$  then shows that

$$\tilde{\mathbb{P}}_\pi(\mathbb{1}_B \circ \tilde{S} = \mathbb{1}_A \circ \tilde{S} = \mathbb{1}_A = \mathbb{1}_B) = 1 .$$

Now, note that  $\mathcal{F}_0 = \sigma(p)$  where  $p$  is defined in (A.4). Then, since  $B \in \mathcal{F}_0$ , there exists  $C \in \mathcal{X}^{\otimes \mathbb{N}}$  such that  $B = p^{-1}(C)$  and thus,

$$\begin{aligned} 1 &= \tilde{\mathbb{P}}_\pi(\mathbb{1}_B = \mathbb{1}_B \circ \tilde{S}) = \tilde{\mathbb{P}}_\pi(\mathbb{1}\{p(\cdot) \in C\} = \mathbb{1}\{p \circ \tilde{S}(\cdot) \in C\}) \\ &= \tilde{\mathbb{P}}_\pi(\mathbb{1}_C \circ p = \mathbb{1}_C \circ p \circ \tilde{S}) \\ &\stackrel{(i)}{=} \tilde{\mathbb{P}}_\pi(\mathbb{1}_C \circ p = \mathbb{1}_C \circ S \circ p) = \tilde{\mathbb{P}}_\pi \circ p^{-1}(\mathbb{1}_C = \mathbb{1}_C \circ S) \stackrel{(ii)}{=} \mathbb{P}_\pi(\mathbb{1}_C = \mathbb{1}_C \circ S) , \end{aligned}$$

where  $\stackrel{(i)}{=}$  follows from  $p \circ \tilde{S} = S \circ p$  and  $\stackrel{(ii)}{=}$  from  $\mathbb{P}_\pi = \tilde{\mathbb{P}}_\pi \circ p^{-1}$  (see (A.3)). The dynamical system  $(X^{\mathbb{N}}, \mathcal{X}^{\otimes \mathbb{N}}, \mathbb{P}_\pi, S)$  being ergodic, it implies that  $\mathbb{P}_\pi(C) = 0$  or 1 which concludes the proof since

$$\mathbb{P}_\pi(C) = \tilde{\mathbb{P}}_\pi \circ p^{-1}(C) = \tilde{\mathbb{P}}_\pi(B) = \tilde{\mathbb{P}}_\pi(A) .$$

□

**Proposition 34.** *Let  $(X^{\mathbb{Z}}, \mathcal{X}^{\otimes \mathbb{Z}}, \mathbb{P}, S)$  be a measure-preserving dynamical system. Then, the following statements are equivalent:*

- (a)  $(X^{\mathbb{Z}}, \mathcal{X}^{\otimes \mathbb{Z}}, \mathbb{P}, S)$  is ergodic.
- (b) for all measurable function  $h : X^{\mathbb{Z}} \rightarrow \mathbb{R}$  satisfying  $\mathbb{E}(h_+) < \infty$ ,

$$n^{-1} \sum_{k=0}^{n-1} h \circ S^k \rightarrow_{n \rightarrow \infty} \mathbb{E}(h), \quad \mathbb{P}\text{-a.s.} \quad (\text{A.9})$$

*Proof.* We first show that (a) implies (b). Assume that  $\mathbb{E}(h_+) < \infty$ . If  $\mathbb{E}(h_-) < \infty$ , then, (A.9) follows from Birkhoff's ergodic theorem. If  $\mathbb{E}(h_-) = \infty$ , then  $\mathbb{E}(h) = -\infty$ . Moreover, since for all nonnegative real number  $M$ ,

$$-M \leq h \mathbb{1}\{h > -M\} \leq h_+ ,$$

the monotone convergence theorem applied to the nondecreasing and nonnegative function,  $h_+ - h\mathbb{1}\{h > -M\}$  yields

$$\lim_{M \rightarrow \infty} \mathbb{E}(h\mathbb{1}\{h > -M\}) = \mathbb{E}(\lim_{M \rightarrow \infty} h\mathbb{1}\{h > -M\}) = \mathbb{E}(h) = -\infty, \quad \mathbb{P}\text{-a.s.}$$

so that  $\mathbb{E}(h\mathbb{1}\{h > -M\}) \rightarrow_{M \rightarrow \infty} -\infty$ . The proof follows from

$$\begin{aligned} \limsup_{n \rightarrow \infty} n^{-1} \sum_{k=0}^{n-1} h \circ S^k \\ \leq \limsup_{n \rightarrow \infty} n^{-1} \sum_{k=0}^{n-1} h \circ S^k \mathbb{1}(h \circ S^k > -M) = \mathbb{E}(h\mathbb{1}\{h > -M\}), \end{aligned}$$

by letting  $M$  goes to infinity. Conversely, assume (b). Let  $A \in \mathcal{X}^{\otimes \mathbb{Z}}$  such that  $\mathbb{1}_A \circ S = \mathbb{1}_A$ . Then,

$$n^{-1} \sum_{k=0}^{n-1} \mathbb{1}_A \circ S^k \rightarrow_{n \rightarrow \infty} \mathbb{P}(A), \quad \mathbb{P}\text{-a.s.}$$

which implies, since  $\mathbb{P}\text{-a.s.}, \mathbb{1}_A \circ S^k = \mathbb{1}_A$ ,

$$\mathbb{1}_A = \mathbb{P}(A), \quad \mathbb{P}\text{-a.s.}$$

Since  $\mathbb{1}_A$  takes value in  $\{0, 1\}$ , then necessarily  $\mathbb{P}(A) = 0$  or  $1$ . The proof is concluded.  $\square$

## Appendix B. Consistency of Max-estimators using stationary approximations

Let  $\mathbf{X}$  be a Polish space equipped with its Borel sigma-field  $\mathcal{X}$  and let  $S$  the shift operator as defined in (A.1). Assume that  $(\mathbf{X}^{\mathbb{Z}}, \mathcal{X}^{\otimes \mathbb{Z}}, \mathbb{P}, S)$  is a measure-preserving ergodic dynamical system. Denote by  $\mathbb{E}$  the expectation operator associated to  $\mathbb{P}$ .

Let  $(\bar{\ell}^\theta, \theta \in \Theta)$  be a family of measurable functions  $\bar{\ell}^\theta : \mathbf{X}^{\mathbb{Z}} \rightarrow \mathbb{R}$ , indexed by  $\theta \in \Theta$  where  $(\Theta, d)$  is a compact metric space and denote  $\bar{\mathbf{L}}_n^\theta := n^{-1} \sum_{k=0}^{n-1} \bar{\ell}^\theta \circ S^k$ . Moreover, consider  $(\mathbf{L}_n^\theta, n \in \mathbb{N}^*, \theta \in \Theta)$  a family of upper-semicontinuous functions  $\mathbf{L}_n^\theta : \mathbf{X}^{\mathbb{Z}} \rightarrow \mathbb{R}$  indexed by  $n \in \mathbb{N}^*$  and  $\theta \in \Theta$ . Consider the following assumptions:

(C3)  $\mathbb{E}(\sup_{\theta \in \Theta} \bar{\ell}_+^\theta) < \infty,$

(C4)  $\mathbb{P}\text{-a.s.},$  the function  $\theta \mapsto \bar{\ell}^\theta$  is upper-semicontinuous,

(C5)  $\lim_{n \rightarrow \infty} \sup_{\theta \in \Theta} |\mathbf{L}_n^\theta - \bar{\mathbf{L}}_n^\theta| = 0, \quad \mathbb{P}\text{-a.s.}$

Let  $\{\bar{\theta}_n : n \in \mathbb{N}^*\} \subset \Theta$  and  $\{\theta_n : n \in \mathbb{N}^*\} \subset \Theta$  such that for all  $n \geq 1$ ,

$$\bar{\theta}_n \in \operatorname{argmax}_{\theta \in \Theta} \bar{\mathbf{L}}_n^\theta, \quad \theta_n \in \operatorname{argmax}_{\theta \in \Theta} \mathbf{L}_n^\theta.$$

Assumptions (C3-4) are quite standard and can be adapted directly from Pfanzagl (1969) (which treated the case of independent  $\{X_n\}_{n \in \mathbb{N}}$ ). For the sake of clarity, we provide here a short and self-contained proof.

**Theorem 35.** *Assume (C3-4).*

- (i) *Then,  $\lim_{n \rightarrow \infty} \mathbf{d}(\bar{\theta}_n, \Theta_\star) = 0$ ,  $\mathbb{P}$ -a.s. where  $\Theta_\star := \operatorname{argmax}_{\theta \in \Theta} \mathbb{E}(\bar{\ell}^\theta)$ .*
- (ii) *Assume in addition that (C5) holds. Then,  $\lim_{n \rightarrow \infty} \mathbf{d}(\theta_n, \Theta_\star) = 0$ ,  $\mathbb{P}$ -a.s. Moreover,*

$$\lim_{n \rightarrow \infty} \mathbf{L}_n^{\theta_n} = \sup_{\theta \in \Theta} \mathbb{E}(\bar{\ell}^\theta), \quad \mathbb{P}\text{-a.s.} \quad (\text{B.1})$$

$$\forall \theta \in \Theta, \quad \lim_{n \rightarrow \infty} \mathbf{L}_n^\theta = \mathbb{E}(\bar{\ell}^\theta), \quad \mathbb{P}\text{-a.s.} \quad (\text{B.2})$$

*Proof. **Proof of** (i).* First note that according to Proposition 34 and (C3), for all  $\theta \in \Theta$ ,  $\lim_{n \rightarrow \infty} \bar{\mathbf{L}}_n^\theta$  exists  $\mathbb{P}$ -a.s., and

$$\lim_{n \rightarrow \infty} \bar{\mathbf{L}}_n^\theta = \lim_{n \rightarrow \infty} n^{-1} \sum_{k=0}^{n-1} \bar{\ell}^\theta \circ S^k = \mathbb{E}(\bar{\ell}^\theta), \quad \mathbb{P}\text{-a.s.} \quad (\text{B.3})$$

Let  $K$  be a compact subset of  $\Theta$ . For all  $\theta_0 \in K$ ,  $\mathbb{P}$ -a.s.,

$$\begin{aligned} & \limsup_{\rho \rightarrow 0} \limsup_{n \rightarrow \infty} \sup_{\theta \in \mathbf{B}(\theta_0, \rho)} n^{-1} \sum_{k=0}^{n-1} \bar{\ell}^\theta \circ S^k \\ & \leq \limsup_{\rho \rightarrow 0} \limsup_{n \rightarrow \infty} n^{-1} \sum_{k=0}^{n-1} \sup_{\theta \in \mathbf{B}(\theta_0, \rho)} \bar{\ell}^\theta \circ S^k = \limsup_{\rho \rightarrow 0} \mathbb{E} \left( \sup_{\theta \in \mathbf{B}(\theta_0, \rho)} \bar{\ell}^\theta \right), \end{aligned} \quad (\text{B.4})$$

where the last equality follows from (C3) and Proposition 34. Moreover, by the monotone convergence theorem applied to the nonincreasing function  $\rho \mapsto \sup_{\theta \in \mathbf{B}(\theta_0, \rho)} \bar{\ell}^\theta$ , we have

$$\limsup_{\rho \rightarrow 0} \mathbb{E} \left( \sup_{\theta \in \mathbf{B}(\theta_0, \rho)} \bar{\ell}^\theta \right) = \mathbb{E} \left( \limsup_{\rho \rightarrow 0} \sup_{\theta \in \mathbf{B}(\theta_0, \rho)} \bar{\ell}^\theta \right) \leq \mathbb{E}(\bar{\ell}^{\theta_0}), \quad (\text{B.5})$$

where the last inequality follows from (C4). Combining (B.4) and (B.5), we obtain that for all  $\eta > 0$  and  $\theta_0 \in K$ , there exists  $\rho^{\theta_0} > 0$  satisfying

$$\limsup_{n \rightarrow \infty} \sup_{\theta \in \mathbf{B}(\theta_0, \rho^{\theta_0})} n^{-1} \sum_{k=0}^{n-1} \bar{\ell}^\theta \circ S^k \leq \mathbb{E}(\bar{\ell}^{\theta_0}) + \eta \leq \sup_{\theta \in K} \mathbb{E}(\bar{\ell}^\theta) + \eta, \quad \mathbb{P}\text{-a.s.}$$

Since  $K$  is a compact subset of  $\Theta$ , we can extract a finite subcover of  $K$  from  $\bigcup_{\theta_0 \in K} \mathcal{B}(\theta_0, \rho^{\theta_0})$ , so that

$$\limsup_{n \rightarrow \infty} \sup_{\theta \in K} n^{-1} \sum_{k=0}^{n-1} \bar{\ell}^\theta \circ S^k \leq \sup_{\theta \in K} \mathbb{E}(\bar{\ell}^\theta) + \eta, \quad \mathbb{P}\text{-a.s.} \quad (\text{B.6})$$

Since  $\eta$  is arbitrary, we obtain

$$\limsup_{n \rightarrow \infty} \sup_{\theta \in K} n^{-1} \sum_{k=0}^{n-1} \bar{\ell}^\theta \circ S^k \leq \sup_{\theta \in K} \mathbb{E}(\bar{\ell}^\theta), \quad \mathbb{P}\text{-a.s.} \quad (\text{B.7})$$

Moreover,  $\mathbb{P}$ -a.s., by (B.5), we get

$$\limsup_{\rho \rightarrow 0} \sup_{\theta \in \mathcal{B}(\theta_0, \rho)} \mathbb{E}(\bar{\ell}^\theta) \leq \limsup_{\rho \rightarrow 0} \mathbb{E} \left( \sup_{\theta \in \mathcal{B}(\theta_0, \rho)} \bar{\ell}^\theta \right) \leq \mathbb{E}(\bar{\ell}^{\theta_0}),$$

This shows that  $\theta \mapsto \mathbb{E}(\bar{\ell}^\theta)$  is upper-semicontinuous. As a consequence,  $\Theta_\star := \operatorname{argmax}_{\theta \in \Theta} \mathbb{E}(\bar{\ell}^\theta)$  is a closed and nonempty subset of  $\Theta$  and therefore, for all  $\epsilon > 0$ ,  $K_\epsilon := \{\theta \in \Theta; d(\theta, \Theta_\star) \geq \epsilon\}$  is a compact subset of  $\Theta$ . Using again the upper-semicontinuity of  $\theta \mapsto \mathbb{E}(\bar{\ell}^\theta)$ , there exists  $\theta_\epsilon \in K_\epsilon$  such that for all  $\theta_\star \in \Theta_\star$ ,

$$\sup_{\theta \in K_\epsilon} \mathbb{E}(\bar{\ell}^\theta) = \mathbb{E}(\bar{\ell}^{\theta_\epsilon}) < \mathbb{E}(\bar{\ell}^{\theta_\star}).$$

Finally, combining this inequality with (B.7), we obtain that  $\mathbb{P}$ -a.s.,

$$\begin{aligned} \limsup_{n \rightarrow \infty} \sup_{\theta \in K_\epsilon} \bar{\mathcal{L}}_n^\theta &= \limsup_{n \rightarrow \infty} \sup_{\theta \in K_\epsilon} n^{-1} \sum_{k=0}^{n-1} \bar{\ell}^\theta \circ S^k \leq \sup_{\theta \in K_\epsilon} \mathbb{E}(\bar{\ell}^\theta) \\ &< \mathbb{E}(\bar{\ell}^{\theta_\star}) \stackrel{(1)}{=} \lim_{n \rightarrow \infty} \bar{\mathcal{L}}_n^{\theta_\star} \leq \liminf_{n \rightarrow \infty} \bar{\mathcal{L}}_n^{\bar{\theta}_n}, \end{aligned} \quad (\text{B.8})$$

where (1) follows from (B.3). This inequality ensures that  $\bar{\theta}_n \notin K_\epsilon$  for all  $n$  larger to some  $\mathbb{P}$ -a.s. finite integer-valued random variable. This completes the proof of (i) since  $\epsilon$  is arbitrary.  $\blacktriangleleft$

**Proof of (ii).** First note that (B.2) follows from (B.3) and (C5).

Let  $\theta_\star$  be any point in  $\Theta_\star$ . Then,  $\mathbb{P}$ -a.s.,

$$\begin{aligned} \mathbb{E}(\bar{\ell}^{\theta_\star}) &\stackrel{(1)}{=} \liminf_{n \rightarrow \infty} \bar{\mathcal{L}}_n^{\theta_\star} \stackrel{(2)}{\leq} \liminf_{n \rightarrow \infty} \bar{\mathcal{L}}_n^{\bar{\theta}_n} \leq \limsup_{n \rightarrow \infty} \bar{\mathcal{L}}_n^{\bar{\theta}_n} \\ &= \limsup_{n \rightarrow \infty} \sup_{\theta \in \Theta} \bar{\mathcal{L}}_n^\theta \stackrel{(3)}{\leq} \sup_{\theta \in \Theta} \mathbb{E}(\bar{\ell}^\theta) = \mathbb{E}(\bar{\ell}^{\theta_\star}), \end{aligned}$$

where (1) follows from (B.3), (2) is direct from the definition of  $\bar{\theta}_n$  and (3) is obtained by applying (B.7) with  $K = \Theta$ . Thus,

$$\bar{\mathcal{L}}_n^{\bar{\theta}_n} \rightarrow_{n \rightarrow \infty} \mathbb{E}(\bar{\ell}^{\theta_\star}), \quad \mathbb{P}\text{-a.s.} \quad (\text{B.9})$$

Denote  $\delta_n := \sup_{\theta \in \Theta} |\mathbb{L}_n^\theta - \bar{\mathbb{L}}_n^\theta|$ . We get

$$\bar{\mathbb{L}}_n^{\bar{\theta}_n} - \delta_n \stackrel{(1)}{\leq} \bar{\mathbb{L}}_n^{\bar{\theta}_n} \stackrel{(2)}{\leq} \mathbb{L}_n^{\theta_n} \stackrel{(1)}{\leq} \bar{\mathbb{L}}_n^{\theta_n} + \delta_n \stackrel{(3)}{\leq} \bar{\mathbb{L}}_n^{\bar{\theta}_n} + \delta_n. \quad (\text{B.10})$$

where (1) follows from the definition of  $\delta_n$ , (2) from the definition of  $\theta_n$  and (3) from the definition of  $\bar{\theta}_n$ . Combining the above inequalities with (B.9) and (C5) yields (B.1). (B.10) also implies that

$$\bar{\mathbb{L}}_n^{\theta_n} \rightarrow_{n \rightarrow \infty} \mathbb{E}(\bar{\ell}^{\theta_*}), \quad \mathbb{P}\text{-a.s.}$$

which yields, using (B.8),

$$\limsup_{n \rightarrow \infty} \sup_{\theta \in K_\epsilon} \bar{\mathbb{L}}_n^\theta < \liminf_{n \rightarrow \infty} \bar{\mathbb{L}}_n^{\theta_n} = \limsup_{n \rightarrow \infty} \bar{\mathbb{L}}_n^{\theta_n} = \mathbb{E}(\bar{\ell}^{\theta_*}), \quad \mathbb{P}\text{-a.s.}$$

where  $K_\epsilon := \{\theta \in \Theta; d(\theta, \Theta_*) \geq \epsilon\}$ . Therefore,  $\theta_n \notin K_\epsilon$  for all  $n$  larger to some  $\mathbb{P}$ -a.s.-finite integer-valued random variable. The proof is completed since  $\epsilon$  is arbitrary.  $\blacktriangleleft$   
 $\square$

**Lemma 36.** *Let  $\{V_n\}_{n \in \mathbb{N}}$  be a sequence of strict-sense stationary random variables on the same probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ . Denote by  $\mathbb{E}$  the associated expectation operator and assume that  $\mathbb{E}[(\ln |V_0|)_+] < \infty$ . Then, for all  $\eta \in (0, 1)$ ,*

$$\lim_{k \rightarrow \infty} \eta^k V_k = 0, \quad \mathbb{P}\text{-a.s.}$$

*Proof.* Let  $\eta \in (0, 1)$ . For all  $\epsilon > 0$ ,

$$\begin{aligned} \sum_{k=1}^{\infty} \mathbb{P}(\eta^k |V_k| \geq \epsilon) &= \sum_{k=1}^{\infty} \mathbb{P}(\ln |V_0| - \ln \epsilon \geq -k \ln \eta) \\ &\leq \sum_{k=1}^{\infty} \mathbb{P}\left(\frac{(\ln(V_0))_+ - \ln \epsilon}{-\ln \eta} \geq k\right) < \infty, \end{aligned}$$

where the last inequality follows from  $\mathbb{E}[(\ln |V_0|)_+] < \infty$ . The proof follows by applying the Borel-Cantelli lemma.  $\square$

## References

## References

- Bougerol, P., Picard, N., 1992. Stationarity of garch processes and of some nonnegative time series. J. Econometrics 52 (1992), 115 – 127.
- Brandt, A., MAR 1986. The stochastic equation  $Y_n = A_n Y_n + B_n$  with stationary coefficients. Adv. Appl. Probab. 18 (1), 211–220.
- Cox, D., 1981. Statistical analysis of time-series: some recent developments. Scand. J. Statist. 8 (2), 93–115.



- Davis, R., Dunsmuir, W., Streett, S., DEC 2003. Observation-driven models for Poisson counts. *Biometrika* 90 (4), 777–790.
- Davis, R., Liu, H., 2012. Theory and inference for a class of observation-driven models with application to time series of counts. Preprint, arXiv:1204.3915.
- Diaconis, P., Freedman, D., 1999. Iterated random functions. *SIAM Rev.* 47 (1), 45–76.
- Doukhan, P., 2012. Comments on: Some recent theory for autoregressive count time series DISCUSSION. *TEST* 21 (3), 447–450.
- Doukhan, P., Fokianos, K., Tjøstheim, D., 2012. On weak dependence conditions for Poisson autoregressions. *Statist. Probab. Lett.* 82 (5), 942–948.
- Doukhan, P., Wintenberger, O., 2008. Weakly dependent chains with infinite memory. *Stochastic Processes and their Applications* 118 (11), 1997 – 2013.
- Dudley, R. M., 2002. *Real Analysis and Probability*. Cambridge University Press.
- Ferland, R., Latour, A., Oraichi, D., 2006. Integer-valued GARCH process. *J. Time Ser. Anal.* 27 (6), 923–942.
- Fokianos, K., Rahbek, A., Tjøstheim, D., 2009. Poisson autoregression. *J. Am. Statist. Assoc.* 104 (488), 1430–1439, with electronic supplementary materials available online.  
URL <http://dx.doi.org/10.1198/jasa.2009.tm08270>
- Fokianos, K., Tjøstheim, D., 2011. Log-linear poisson autoregression. *J. of Multivariate Analysis* 102 (3), 563–578.
- Gray, R., 2009. *Probability, Random Processes, and Ergodic Properties*. Springer, London.
- Hairer, M., Mattingly, J., 2006. Ergodicity of the 2d navier-stokes equations with degenerate stochastic forcings. *Ann. Math.* 164, 993–1032.
- Hall, P., Heyde, C. C., 1980. *Martingale Limit Theory and its Applications*. Academic Press, New York, London.
- Henderson, S. G., Matteson, D., Woodard, D., 2011. Stationarity of generalized autoregressive moving average models. *Electronic Journal of Statistics* 5, 800–828.
- Kedem, B., Fokianos, K., 2002. *Regression models for time series analysis*. Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ.
- Meyn, S. P., Tweedie, R. L., 1993. *Markov Chains and Stochastic Stability*. Springer, London.

- Neumann, M. H., NOV 2011. Absolute regularity and ergodicity of Poisson count processes. *Bernoulli* 17 (4), 1268–1284.
- Pfanzagl, J., 1969. On the measurability and consistency of minimum contrast estimates. *Metrika* 14, 249–272.
- Streett, S., 2000. Some observation driven models for time series of counts. Ph.D. thesis, Colorado State University, Department of Statistics.
- Sunyach, C., 1975. Une classe de chaînes récurrentes sur un espace métrique complet. *Ann. Inst. H. Poincaré Probab. Statist.* XI, 325–343.
- Tjøstheim, D., 2012. Some recent theory for autoregressive count time series. *TEST* 21 (3), 413–438.
- Tweedie, R. L., 1988. Invariant measures for Markov chains with no irreducibility assumptions. *Journal of Applied Probability* 25, 275–285.
- Zhu, F., 2012. Modeling overdispersed or underdispersed count data with generalized Poisson integer-valued GARCH models. *J. of Mathematical Analysis and Applications* 389 (1), 58–71.